Dipartimento di Fisica "E. Fermi" / **Università di Pisa**

COMPLEX PROJECTIVE MODEL ON A WORLDSTRIP

TESI DI LAUREA MAGITRALE IN FISICA

Relatore
Chiar.mo Prof. KENICHI KONISHI

Candidato
ALESSANDRO BETTI

Anno Accademico 2015/2016

PREFACE

To doubt everything or to believe everything
are two equally convenient solutions;
both dispense with the necessity of reflection.
— JULES HENRI POINCARÉ, quoted by Bertrand Russel (1913)

THE TOPICS that we will discuss in this work are all intended to give the necessary background for the understanding of a model, the Complex Projective model defined on a worldstrip, that will be the main topic of Chapter 3.

The title "Complex Projective model on a Worldstrip" may sound strange and unusual at first reading. The system we are ultimately interested in and the one the title refers to is the well known $\mathbf{C}\mathrm{P}^{N-1}$. The term "worldstrip" is used here instead of the more common "worldsheet" since the word "sheet" suggests an infinite extension, while we want to stress the fact that we are going to take as a domain for our fields a spacially bounded strip.

Beside the inherent difficulty of studying a quantum field theory model on a bounded region of space (due to the existence of boundary effects), the nature of the problem and the way in which it emerges into the panorama of modern theoretical physics, forces us to take into account not only the standard techniques and ideas of ordinary quantum field theory but also those developed in the area of topological solitons.

The thesis is organized in *three* chapters: In Chapter 1, after a general introduction to the problem, we go through some mathematical preliminaries in preparation for the second part of the chapter in which we review the basics of soliton theory. In this chapter there is a high percentage of mathematical material that in our opinion cannot be avoided if one wants to have a good understanding of the subject. The chapter ends with a section that explains how a theory that possesses a nonabelian vortex confined by nonabelian monopoles leads to the study of $\mathbf{C}\mathrm{P}^{N-1}$ defined on a spatially bounded spacetime worldstrip.

Chapter 2 gives an introduction and a review of one of the tools that we will use in the study of the model: The large N expansion. Topological discussions involving the Euler characteristic are formalized using the mathematical concept of CW complexes. Such technique is applied, eventually, to the resolution of the $\mathbf{C}\mathrm{P}^{N-1}$ model on the plane. Essential properties such as asymptotic freedom and confinement are discussed.

The main part of the work is contained in Chapter 3 in which the complex projective model is studied on the worldstrip. First of all the effective action and the full gap equations are derived; the problem of boundary conditions

is addressed and discussed in the case of Dirichlet and Neumann boundary conditions and in the case of periodic conditions. Then the gap equations are analyzed and it is shown that translationally invariant solutions are compatible only with periodic conditions. The numerical analysis in the case of general boundary conditions is discussed.

In the last section of the chapter we study the energy of the model. We recover the results known in the literature for periodic boundary condition. The energy density for a theory with general boundary conditions is derived and we subsequently prove that it is finite after an appropriate subtraction. We manage to show that from the expression derived for the energy density it is indeed possible to recover the generalized gap equations, providing us with an important consistency check.

Finally it is proven that the energy density is actually constant and we end with some considerations on the large L behavior of the energy, where L is the spatial width of the worldstrip.

In Chapter 1 and 2 we tried to give a detailed and exhaustive discussion of the various topics there treated but at the same time avoiding technicalities for which we directly refer to the literature. On the other hand Chapter 3, since it contains also original work, goes more deeply into the analysis and offers more exhaustive calculations.

A final remark: Particular care has been put into the design and realization of figures, most of them have been produced with John Hobby's METAPOST system, the others have been prepared with *Adobe Illustrator*.

Pisa A.B.
September 2016

I have no data yet.
It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts.

— Sherlock Holmes, in *A Scandal in Bohemia*, (1891)

You see, but you do not observe.
The distinction is clear.

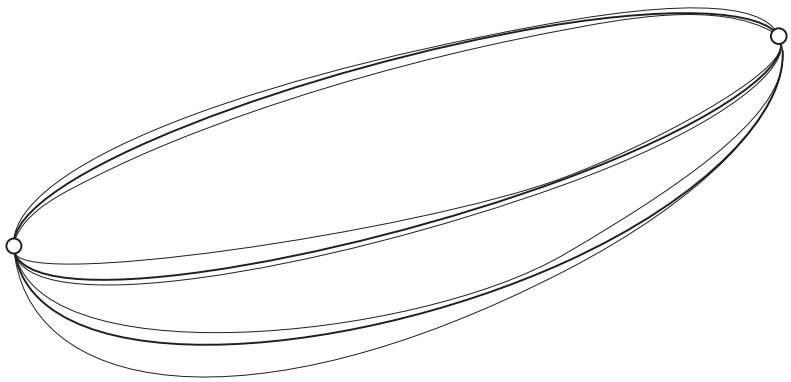
— Sherlock Holmes, in *A Scandal in Bohemia*, (1891)

We can face our problem.
We can arrange such facts as we have
with order and method.

— HERCULE POIROT, in murder on the Orient Express (1934)

CONTENTS

Thou art the Mars of malcontents.
— WILLIAM SHAKESPEARE, *The Merry Wives of Windsor* (I, iii, 111)



CHAPTER ONE

BASIC CONCEPTS

For this chapter,
it will suffice to provide a basic understanding of the tool and its technique.
Readers are encouraged to take this chapter not as an exhaustive reckoning,
but as a firm foundation upon which to build edifices of their own experience.

— DAVID REES, *How to Sharpen Pencils* (2012)

Wait! Now, you see, the thing is very simple;
this time you'll understand it.

— MARK TWAIN, *A Connecticut Yankee in King Arthur's Court* (1889)

"Can it matter?" He asked.
"It is most essential."

— SHERLOCK HOLMES, in *The Adventure of the Blanched Soldier* (1926)

THE MODEL we will discuss in this thesis, the $\mathbf{C}\mathrm{P}^{N-1}$ model, has — as many other theories in QFT — an interesting historical development. It was first introduced in the late seventies (around 1978–1979) as an example of non-linear $SU(N)$ sigma model by Harald Eichenherr [26], Voislav L. Golo and Askold M. Perelomov [27] and by Eugène Cremmer and Joël Scherk in [19]. Then it was analyzed in the large N limit by Alessandro D'Adda, Paolo di Vecchia, Martin Lüscher (see [20] and [21]) and in a separate article in the same year by Edward Witten [56]; these people studied the system in order to improve their understanding of QCD. In particular in Witten's article it was pointed out how that model shares lots of common features with QCD. Following this spirit we will investigate the model in a "traditional way " in Chapter 2.

Yet there is another reason that motivates us to pursue the study of this model with renewed interest. This reason is that such model emerges naturally in the study of some solitonic solutions thought to be relevant for the problem of confinement in QCD. This line of research has expanded after the so called nonabelian solitonic solutions of vortices and monopoles have been discovered in the early 2000s. Since then many research groups have begun to study the model, because within this framework there is the necessity of studying the problem on a spatially bounded region. The study of the matter under this new light is proposed in Chapter 3, and as we will see there are many aspects of the problem to be fully understood yet.

While the technical tools to address the problem in its original formulation are described in Chapter 2, in a self contained way, we feel that in order to

fully appreciate the material presented in Chapter 3 we need to go through the basics of solitons and this in turns requires a basic understanding of some mathematical concepts, especially of topology and algebraic topology. This is exactly the purpose of this chapter, that will end with the presentation of a QFT model that will make how $\mathbf{C}\mathrm{P}^{N-1}$ enters in this story clear.

Before going into the study of some mathematical preliminaries let us briefly explain why soliton-like solutions, are candidates to shed some light on confinement in QCD.

Many efforts have been put by researchers into the systematic study of vortices and monopoles because it is believed that they might play an important role in the explanation of confinement in QCD. The basic idea has been proposed by 't Hooft [52] and Mandelstam, [35] and [36], who conjectured that the ground state of QCD is a dual superconductor where quarks are confined by chromoelectric ANO vortices, that dualize what happens in type II superconductors where magnets are confined by a magnetic field. (Later in this chapter we will carefully explain what we mean by a string or a vortex, for the moment just think of these objects as classical field configurations.)

Such mechanism has in fact been studied in supersymmetric theories; for example in the work [47] by Seiberg and Witten a $\mathcal{N} = 2$ supersymmetric theory with gauge group $SU(2)$ has been studied and they found two singularities in the vacuum of the theory that they interpret as massless monopoles.

Originally the solution proposed by 't Hooft and Mandelstam was abelian. Assuming such an abelian model if one wants to start from a general $SU(N)$ gauge group then he has to assume that the system abelianize to $(U(1))^{N-1}$, because otherwise at low energy the system could not be described in terms of abelian monopoles.

This abelian description, however, have little to share with the actual QCD (first of all the spectrum of the theory is different), and then people started to realize that it could have been more natural to consider nonabelian degrees of freedom; from this kind of observations the study of nonabelian vortices and monopoles took the first moves.

So what one has to look for are models that show a mechanism similar to the one that describes the Meissner effect in superconductors but that involve nonabelian objects in such a way that there might be hope for such system to correctly describe the ground state of QCD.

The subject and the number of papers that appeared in the literature in the last years is enormous, and a lot of subtle effects that enter in a non-trivial way in the discussion should be described in order to have a satisfactory picture of the problem (probably because we still lack of a good comprehension of the problem). All this analysis would require a treatment on their own also because they cannot be understood without a good knowledge of supersymmetric theories; all this is beyond the scope of this work.

The main idea however is that one can find models in which nonabelian monopoles are confined by nonabelian vortices in a monopole-vortex-monopole (M-V-M) configuration. This kind of systems gives us the opportunity to study

the vacuum of QCD (r-vacua, see for example [5]) from a semiclassical, and more intuitive point of view. Moreover they give us the motivation for the study of the $\mathbf{C}\mathrm{P}^{N-1}$ model on a finite worldstrip since it will emerge as a worldsheet theory of the M-V-M solution as it is discussed in more details in Section 1.2.4. An excellent review that covers all these topics is [31].

A last comment before we begin with the various discussions; in the models that we will consider we will always have, next to a color symmetry, a flavor symmetry. This is not an accident but it is something that is essential for the realization of the M-V-M complex as it is discussed in details in [32].

Topology is the property of something that doesn't change when you bend it or stretch it as long as you don't break anything.

— EDWARD WITTEN, NOVA interview (2003)

1.1. MATHEMATICAL PRELIMINARIES

In what follows we would like to give a self contained discussion about some of the mathematical techniques that we will use later in this chapter. We will try to keep the discussion rigorous from a mathematical point of view, but at the same time to give some visual and practical insight of the topic under discussion.

The main result that we will review are those connected with algebraic topology. We focus on those contents mainly because they don't have a standard presentation in the curriculum of a student in Physics and therefore we think that some discussions are needed at least to fix the notations.

We will also need as a prerequisite some results from point-set topology. Because this is a highly standardized subject we will only limit ourselves to recall some of the main concepts and terminology for further use, this time trying to really furnish some intuition rather that a polished set of definitions and propositions. And from intuition we will try to make sense of some of the well known facts that concern topology.

Finally we will also need to use some results from the theory of integration over smooth manifolds in order to define what is known as the degree of a map.

The main reference that we will use for algebraic topology is the marvelous book from Allen Hatcher *Algebraic Topology* [29], while for further references on the subject of smooth manifolds we refer to the book *Introduction to Smooth Manifolds* by John Lee [34]. Both this texts are quite recent and cover the respective subjects in detail. Most of the discussion here given are transpositions from the material covered in these books.

1.1.1. Point-set Topology

First of all some terminology. If one searches for the word *topology* in the *Oxford Dictionary of English* he will discover that there are two possible meaning attached to the word depending on the context in which it is used. The first one refers to the area of mathematical research that studies the "geometrical properties and spatial relations unaffected by the continuous change of shape or size of the figures." The other definition refers to the structures (a set of objects) that one can define on a space to make it suitable for topological studies in the way that we have just mentioned. Let us be more specific about this.

As Allen Hatcher says topology is "qualitative geometry", meaning it is the study of geometrical objects modulo continuous deformations, that is to say objects that can obtained one from the other by bending or stretching but not cutting or breaking or trimming away pieces. It is most like doing an origami but with a sheet of rubber instead of a sheet of paper.

To be precise we have to specify what those geometrical objects are, and what are the transformations, between the allowed ones, that can turn one of this objects into another so that they can be identified. The crucial concept that bonds together all these ideas is the continuity. The topological equivalence is formulated in terms of maps, called *homeomorphism*, that are essentially defined as continuous relations (that do not break the space), between geometrical objects — in this context called topological spaces — that are also are continuous in a sense that will be precised in a moment.

So, what is a topological space? Answer: A topological space is a space for which continuous functions can be defined.

Of course this answer relies on the concept of continuity of a function that is certainly familiar to readers in terms of limits or ϵ, δ statements; however here we want another definition in terms of open sets.

For concreteness let us focus our attention just on functions $f\colon \mathbf{R} \to \mathbf{R}$, and say that f is continuous if for every open set $O \subset \mathbf{R}$ then

$$f^{-1}(O) \subset \mathbf{R} \quad \text{is also open.} \tag{1}$$

This definition is actually the same as the one we are used to, stated in a different way and it actually means that the images of a sequence of points that are getting indefinitely near to a fixed point x_0 eventually will become indefinitely near the image of the point $f(x_0)$.

Let us see this in some details. Suppose that the condition in Eq. (1) does not hold for the function f, then this means that there would exists some O such that $f^{-1}(O)$ is not open. But if $f^{-1}(O)$ is not open it means that there must exist some point x_0 in $f^{-1}(O)$ such that it is not possible to choose an open interval $(a\,..\,b)$ that contains x_0 and it is completely contained in $f^{-1}(O)$. Since O is an open set there are points that are near $f(x_0)$ in the sense that they belong to an open interval $(c\,..\,d)$ completely contained in O. This practically means that there are points arbitrarily close to x_0 with images at fixed distance from $f(x_0)$.

This intuitive definition of topology we have given here, as a property that the spaces must have in order to be able to talk about continuous functions, is actually better formalized in terms of the classical construction with open sets; for details on this standard construction see for example the first chapter of *Real and Complex Analysis* by Walter Rudin (see [46]).

The last concept we would like to recall before going on to talk more about advanced concept of algebraic topology are the *quotient spaces*. We will see this procedure in action in Chapter 2 while talking about Cell Complexes, and as we will remind there quotient spaces are a convenient way to construct complicated spaces out of simpler ones giving some gluing instructions.

The gluing process can be described in terms of a map $f\colon X \to Y$, that is onto. It is clear that it cannot be one-to-one since the "edges" of X that are glued together are sent by f in the same point of Y.

The remarkable feature of all this is that since f is onto there is a theorem that states that any continuous map from a compact space to an Hausdorff space that is onto defines a topology on Y that is called quotient topology.

This fact allow us to describe very complicated spaces in a simple way. For example a torus can be described from a square once the opposite sides are glued together (for examples see Chapter 2).

1.1.2. Homotopy and Fundamental Group

As one can guess, the study of topological spaces is not easy from a mathematical point of view because — as it happens with all problems involving equivalence classes — one has to abstract from the particular form of an example general properties defined unless of equivalence relations. On the other hand in Physics, at least in the study of solitonic solutions, it is essential to be able to characterize topological spaces.

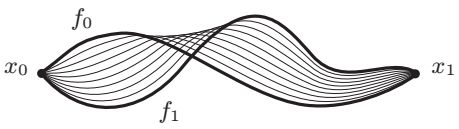
Algebraic topology tries to solve this problem by finding relations between topological spaces and algebraic structures such as groups. This method of investigation comes with an additional feature, that is the fact that not only one is able to obtain algebraic representations of topological spaces, but also of maps between them.

The simplest of such algebraic images, is the fundamental group, the group that one obtains from a topological space X built studying the loops on the space.

This approach to the problem is very much in the spirit of physics methodologies, where the idea is that since you do not have direct access to the structure of the topological space, you probe it trowing loops on it and you see what happens if you rewind the loop keeping one point nailed to the space. Two things can happen, either the loop completely rewinds on the chosen fixed point, or it will remain stuck in and you won't be able to rewind it anymore. In the second case it means that you have discovered a "hole" in the space.

This is exactly the kind of things that the fundamental group is good at: Finding holes of spaces. We could go on and try to describe in an intuitive way why this kind of structure is a group, however at this stage it is probably better to formalize mathematically the concepts we have just tried to explain.

First of all we need a tool that is called *homotopy* and it is exactly the formalization of what earlier we have called "rewinding". Now, the rewinding in mathematics can be described in terms of continuous transformations of a path. Here by *path* in a topological space X we mean a continuous map $f\colon I \to X$, where I is the closed unit interval $[0\,..\,1]$. A *homotopy* of paths in X is a family $f_t\colon I \to X$, with $t \in I$ such that



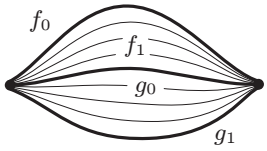
i) for every t the endpoints of the paths are the same: $f_t(0) = x_0$ and $f_t(1) = x_1$,

ii) the associated map $F\colon I \times I \to X$ defined by $F(s,t) \equiv f_t(s)$ is continuous.

For the purpose of defining the fundamental group actually we only need to consider homotopy of closed loops that is to say $x_0 \equiv x_1$. Given two paths f_0 and f_1, if they are connected by an homotopy they are said to be *homotopic* and we write $f_0 \simeq f_1$.

An easy example of homotopy is the linear homotopy in $\mathbf{R}^n$: Given two paths f_0 and f_1 in $\mathbf{R}^n$ with the same endpoints x_0 and x_1 they are homotopic via $f_t(s) = (1-t)f_0(s) + tf_1(s)$.

One important fact about homotopy is that it is an equivalence relation in the sense of being symmetric, reflexive and transitive. Reflexivity is evident since $f \simeq f$ via the constant homotopy $f_t = f$ for all $t \in I$. Symmetry is also easy to prove since if $f_0 \simeq f_1$ via f_t then $f_1 \simeq f_0$ via f_{1-t}. Transitivity follows from the following reasoning: If $f_0 \simeq f_1$ via f_t and $f_1 = g_0$ with $g_0 \simeq g_1$ via g_t then $f_0 \simeq g_1$ via h_t defined to be $f(2s)$ if $0 \leq t \leq 1/2$ and g_{1-2t} if $1/2 \leq t \leq 1$. Continuity of $H(s,t) = h_t(s)$ follow from the fact that a function defined on the union of two closed sets is continuous if it is continuous when restricted to each of the closed sets separately.



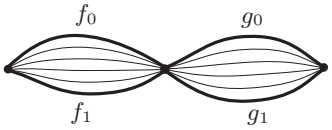
Therefore we have proved that paths that are related by an homotopy can be grouped together into equivalence classes. If f is a path we denote the equivalence class of f as $[f]$ and we call it the homotopy class of f.

The other important fact of homotopy is that there is a natural way to define a multiplication rule between two classes of homotopic loops, this property is the crucial point for the definition of the fundamental group.

Let us start by noticing that given two paths $f, g\colon I \to X$ such that $f(1) = g(0)$ we can define the composition or the product path $f \cdot g$ as

$$(f \cdot g)(s) = \begin{cases} f(2s), & \text{if } 0 \leq s \leq 1/2; \\ g(1-2s), & \text{if } 1/2 \leq s \leq 1. \end{cases} \tag{1}$$

We concatenate the two paths each of which is traversed at double speed. This operation preserves homotopy relations: Suppose that $f_0 \simeq f_1$ via f_t and $g_0 \simeq g_1$ via g_t with $f_0(1) = g_0(0)$ then $f_t \cdot g_t$ is well defined and is exactly the homotopy through which we say that $f_0 \cdot g_0 \simeq f_1 \cdot g_1$.



Now let us consider paths $f\colon I \to X$ with the same starting and ending point $f(0) = f(1) = x_0$, or equivalently paths $f\colon S^1 \to X$, where S^1 is the one dimensional sphere (here we have imposed the constraint $f(0) = f(1)$ gluing the endpoint of I so that this requirement is satisfied by definition), such paths are called loops and x_0 is referred to as the basepoint. The set of all homotopy classes of paths on the space X with basepoint x_0 is denoted by $\pi_1(X, x_0)$.

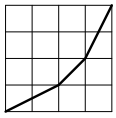
The claim is that this set has a group structure with respect to the product $[f][g] = [f \cdot g]$, and it is called the *fundamental group* of X. (This group does not depend on the particular choice of the basepoint if X is path connected.)

So we would like to prove the following theorem:

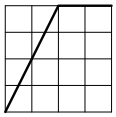
Theorem A. *The fundamental group $\pi_1(X, x_0)$ is a group with respect to the product $[f] \cdot [g] = [f \cdot g]$.*

Proof. First of all let us check that if $[f]$ and $[g]$ are elements of $\pi_1(X, x_0)$, then $[f] \cdot [g]$ is well defined. This is almost obvious since if f and g are loops starting and ending at x_0, then, $f \cdot g$ because of Eq. (1), is a loop starting and ending at x_0. Also, since the homotopy class of $f \cdot g$ is determined uniquely by the homotopy classes of $[f]$ and $[g]$, the multiplication of the group is well defined.

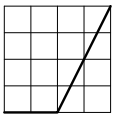
Now we need to check for the defining properties of a group: Associativity, the existence of a neutral element and the existence of the inverse. In order to establish this facts it is extremely useful to use what is called a reparametrization of a path. The basic idea is to change the speed at which the parameter s changes along the path $f(s)$ through a continuous function $\varphi\colon I \to I$ such that $\varphi(0) = 0$ and $\varphi(1) = 1$; in this way the composite function $f\varphi(s) = f(\varphi(s))$ is another path with the same endpoints of f. Moreover a homotopy class is stable under reparametrization, since any reparametrization of a path $f\varphi$ is homotopic to f via the homotopy $f\varphi_t$, where φ_t interpolates between φ and s that is to say $\varphi_t(s) = (1-t)\varphi(s) + ts$.



Now associativity can be easily proved. What we have to prove is that, given three paths f, g and h, with $f(1) = g(0)$ and $g(1) = h(0)$, we have that $([f] \cdot [g]) \cdot [h] = [f] \cdot ([g] \cdot [h])$. This condition is of course satisfied if and only if $(f \cdot g) \cdot h \simeq f \cdot (g \cdot h)$. If one thinks of what Eq. (1) means, it appears obvious that $f \cdot (g \cdot h)$ is just a reparametrization of $(f \cdot g) \cdot h$, and a little thinking shows that the function φ in this case is exactly that in the figure on the side of the page.



Let us now prove that, for each path f, there exists another path c such that $[f]\cdot[c] = [c]\cdot[f] = [f]$. This in practice means that we have to check that there exists a path c such that $f \cdot c \simeq f$ and $c \cdot f \simeq f$. In order to prove the first assertion take c to be the constant path at $f(1)$, that is to say $c(s) = f(1)$ for all $s \in I$, then $f \cdot c$ is just a reparametrization of f once we choose φ to be the function at the side of this page. Similarly one can prove that, if we choose c to be the path such that $c(s) = f(0)$ for all $s \in I$, then $c \cdot f \simeq f$ via the reparametrization specified by the figure at the side. Then, since we are interested in closed loops at x_0, this proves that a constant path at x_0 is the neutral element of $\pi_1(X, x_0)$.



Now given a path f, the inverse path f^{-1} is defined by $f^{-1}(s) = f(1-s)$. With similar arguments to the ones we have used above one can show that in fact $f^{-1} \cdot f$ is homotopic to a constant path, and taking f to be a loop at x_0 it is clear that $[f^{-1}]$ is the inverse of $[f]$ in $\pi_1(X, x_0)$. ∎

One of the most important examples of a space with non trivial fundamental group is the circle S^1, for which we have that $\pi_1(S^1) \approx \mathbf{Z}$. Because of its importance we can state this result as a theorem:

Theorem B. *The group $\pi_1(S^1)$ is an infinite cyclic group generated by the homotopy classes of the loops $(\cos 2\pi s, \sin 2\pi s)$ based at $(1, 0)$.*

A serious mathematical proof of this fact does not come without some work. However intuitively we can understand this result quite easily: The representative of the n-th homotopy class is the loop that starts, for example at $(1, 0)$ and winds n times (clockwise if n is negative and anticlockwise if n is positive around the circle) at constant speed, any other loop can be deformed into one of those loops. On the other hand it is clear that if $n \neq m$ it cannot exist an homotopy that connects the two paths.

Notice also that in Theorem B we have written $\pi_1(S^1)$ without specifying the base point of the space x_0. The reason for this is that it is very easy to prove (see for example [29]), and equally simple to believe that if the space X is path connected all the fundamental groups at different basepoints are isomorphic, then we can safely drop the indication of x_0.

The fundamental group also has a nice property with respect to the direct product, in fact the fundamental group of a product space is isomorphic to the product of the fundamental groups of its factors

$$\pi_1(X \times Y) \approx \pi_1(X) \times \pi_1(Y). \tag{2}$$

This is indeed a quite powerful result since it enables us to compute a lot of fundamental groups without doing any kind of effort. For example the fundamental group of a torus comes for free from Eq. (2). In fact as a topological space the torus T is homeomorphic to $S^1 \times S^1$, and this immediately gives us $\pi_1(T) \approx \mathbf{Z} \times \mathbf{Z}$.

Homotopy groups. The fundamental group is a good tool to investigate topological spaces of low dimensions; however as one can easily understand when one has to study spaces in higher dimensions it will happen that the fundamental group won't be able to give us much information, most of the times it will be trivial. The main reason for this is that the philosophy behind the fundamental group that we have explained in the previous paragraphs — throwing circles on topological spaces (as depicted in Fig. 1) and then see whether they can be reduced to the trivial loop (or more in general see in which homotopy class they end up) — when one has a lot of directions in which you can move is not very effective, because you can unwind the loop on the space making use of those extra dimensions. Basically you are trying to catch structures with a net whose mesh is too big to keep them.

The way to get around this problem is obviously to use a net with smaller meshes, that in our case means that we must throw higher dimensional spheres on our topological spaces.

This idea leads to a natural extension of the fundamental group, the homotopy groups, those structures as we will understand are effective in the study of higher dimensional spaces.

Let us now give a formal definition. Consider an n-dimensional unit cube $I^n = I \times \cdots \times I$ and consider a topological space X with a fixed basepoint x_0, then $\pi_n(X, x_0)$ is defined as the set of homotopy classes of applications $f\colon (I^n, \partial I^n) \to (X, x_0)$ with the constraint that the surface of the unit cube is

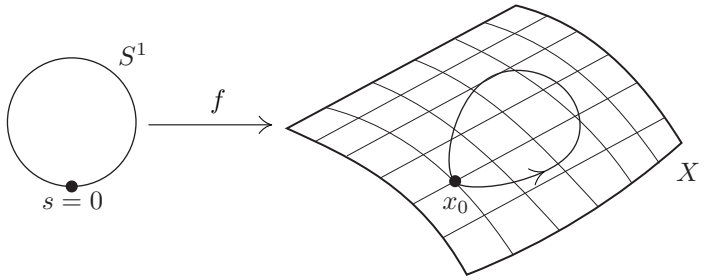


Fig. 1. Mapping of S^1 into X with basepoint $f(0) = x_0$.

mapped into the base point (this is the equivalent of considering closed loops in the definition of the fundamental group), that is $f_t(\partial I^n) = x_0$ for every $t \in I$. Two comments here, on this general definition, are in order. The first one is that of course for $n = 1$ we exactly recover the definition of the fundamental group, the other is that now that we have a definition for a generic n (that make sense for $n \geq 0$) as always in mathematics, it is extremely instructive to look also backwards to the case $n = 0$. In this case, $\pi_0(X, x_0)$ is not a group, however it counts the path-components of X, since you can continuously move a point only within a path-connected portion of a space.

Now let us see what we can say about the group structure of π_n. As we have done before for the fundamental group, when $n \geq 2$, we can define an operation between elements of π_n just by defining a "composition rule" between two maps of the form that we have used to specify the homotopy relation: consider $f, g\colon (I^n, \partial I^n) \to (X, x_0)$ then the composition rule is defined by

$$(f + g)(s_1, s_2, \ldots, s_n) = \begin{cases} f(2s_1, s_2, \ldots, s_n), & \text{if } 0 \leq s_1 \leq 1/2; \\ g(1 - 2s_1, s_2, \ldots, s_n), & \text{if } 1/2 \leq s_1 \leq 1. \end{cases} \tag{3}$$

This rule is well defined on homotopy classes and induces an operation on π_n: $[f] + [g] = [f + g]$. With this operation we can prove that $\pi_n(X, x_0)$ is a group: Since the composition rule involves only the first parameter the proof is the same as for the π_1 case. The identity element of this group is the constant map that send I^n into the basepoint, while the inverse is the equivalence class containing the mapping $-f(s_1, s_2, \ldots, s_n) = f(1 - s_1, s_2, \ldots, s_n)$.

Notice also that we have used for the group operation the sign + instead of the multiplication because for $n \geq 2$, $\pi_n(X, x_0)$ is abelian, that is $[f+g] = [g+f]$.

It is sometimes useful to think of maps $(I^n, \partial I^n) \to (X, x_0)$ as maps from the quotient $I^n/\partial I^n \equiv S^n$ to X taking the basepoint $s_0 = \partial I^n/\partial I^n$ to x_0. This allow us to visualize $\pi_n(X, x_0)$ as homotopy classes of maps $(S^n, s_0) \to (X, x_0)$. This is exactly the kind of intuition that led us to the definition of the homotopy groups: The idea of mapping spheres on topological spaces.

Again if X is path-connected we can prove that different choices of x_0 produce isomorphic homotopy groups, therefore in this cases we can drop the x_0 and write $\pi_n(X)$. Also the behavior of homotopy groups with respect to products is very simple, in fact if we consider a product $\prod_\alpha X_\alpha$ of path connected spaces

Table 1

HOMOTOPY GROUPS FOR SPHERES $\pi_i(S^n)$

n	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$	$i=11$	$i=12$
1	$\mathbf{Z}$	0	0	0	0	0	0	0	0	0	0	0
2	0	$\mathbf{Z}$	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{12}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_3$	$\mathbf{Z}_{15}$	$\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
3	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{12}$	Z_2	$\mathbf{Z}_2$	$\mathbf{Z}_3$	$\mathbf{Z}_{15}$	$\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
4	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}\times\mathbf{Z}_{12}$	$\mathbf{Z}_2\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$	$\mathbf{Z}_{24}\times\mathbf{Z}_3$	$\mathbf{Z}_{15}$	$\mathbf{Z}_2$
5	0	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{24}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{30}$
6	0	0	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{24}$	0	$\mathbf{Z}$	$\mathbf{Z}_2$
7	0	0	0	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{24}$	0	0
8	0	0	0	0	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{24}$	0

This table is taken from the book by Allen Hatcher *Algebraic Topology*, (see [29]). For more extended tables see the book by Hiroshi Toda *Composition methods in homotopy groups of spheres*, Princeton University Press (1962).

X_α, then

$$\pi_n\Big(\prod_\alpha X_\alpha\Big) \approx \prod_\alpha \pi_n(X_\alpha). \tag{4}$$

What we have explained here are the basic definitions and properties of homotopy groups. Now the real goal of algebraic topology is to calculate homotopy groups of all the topological spaces you can think of. Here we cannot go into detailed calculations of the homotopy groups that we will need in what follows because they involve a bunch of technical mathematical tools that would need separate discussion. So we will limit ourselves to give the results of some of the calculations we will use in this chapter. However some of the results can be readily appreciated if one depict a geometrical representation of the problem (as we did for example for the fundamental group of the circle). Homotopy groups $\pi_i(S^n)$ are written in Table 1 up to $i=12$ and $n=8$.

As one can see from Table 1, and as we expected, the fundamental group is useless for higher dimensional spaces (in fact for spheres is trivial for $n\geq 2$). Also $\pi_i(S^n)\approx 0$ if $i<n$, and more interestingly $\pi_n(S^n)\approx\mathbf{Z}$ for $n=1,2,\dots$.

For our purposes, however homotopy groups of spheres are not enough. Since we will have to deal with gauge groups, we also need to know homotopy groups of some Lie groups. Some of this results are listed in Table 2.

Finally we must be also prepared to compute homotopy group of coset spaces. To do this there exists a series of results that connect homotopy groups of different order; for example if G is a Lie group with a subgroup H with G connected and simply connected then we can prove that $\pi_2(G/H)$ and $\pi_1(H)$ are isomorphic as well as $\pi_1(G/H)$ and $\pi_0(H)$.

In order to better explain this results we need to introduce a very useful concept that is that of an exact sequence.

Exact Sequences. A powerful tool that enables many calculations of homotopy groups is what is called an exact sequence.

The definition of an exact sequence is rather simple and at first look seems harmless. However as we will see its strength is that we can express many

Table 2

OTHER COMMONLY USED HOMOTOPY GROUPS FOR SOME LIE GROUPS

Lie Group G	$\pi_1(G)$	$\pi_2(G)$	$\pi_3(G)$	$\pi_4(G)$	$\pi_5(G)$	$\pi_6(G)$
$G = SO(3)$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{12}$
$G = SO(4)$	$\mathbf{Z}_2$	0	$\mathbf{Z} \times \mathbf{Z}$	$\mathbf{Z}_2 \times \mathbf{Z}_2$	$\mathbf{Z}_2 \times \mathbf{Z}_2$	$\mathbf{Z}_{12} \times \mathbf{Z}_{12}$
$G = SO(5)$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0
$G = SO(6)$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0	$\mathbf{Z}$	0
$G = SO(n), \quad n > 6$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0	0	0
$G = U(1)$	$\mathbf{Z}$	0	0	0	0	0
$G = SU(2)$	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_{12}$
$G = SU(3)$	0	0	$\mathbf{Z}$	0	$\mathbf{Z}$	$\mathbf{Z}_6$
$G = SU(n), \quad n > 3$	0	0	$\mathbf{Z}$	0	$\mathbf{Z}$	0

The entries of this table are taken from *Geometry, Topology and Physics* by Mikio Nakahara [42], and from Appendix B of Wienberg's *The Quantum Theory of Fields*, vol. 2 chapter 23 [54].

algebraic concepts in term of exact sequences in a way that is very suitable for homotopy manipulations.

An *exact sequence* is a sequence of homomorphisms α_n between spaces A_n

$$\cdots \longrightarrow A_{n+1} \xrightarrow{\alpha_{n+1}} A_n \xrightarrow{\alpha_n} A_{n-1} \longrightarrow \cdots, \tag{5}$$

such that $\operatorname{Ker} \alpha_n = \operatorname{Im} \alpha_{n+1}$ for each n. With this definition we now show how the concept of injectivity and surjectivity (and therefore isomorphism) of a map between spaces can be expressed in terms of exact sequences. The following facts holds:

i) the sequence $0 \to A \xrightarrow{\alpha} B$ is exact if and only if $\operatorname{Ker} \alpha = 0$ (α is injective);

ii) the sequence $A \xrightarrow{\alpha} B \to 0$ is exact if and only if $\operatorname{Im} \alpha = B$ (α is surjective);

iii) because of i) and ii), the sequence $0 \to A \xrightarrow{\alpha} B \xrightarrow{\beta} 0$ is exact if and only if α is an isomorphism;

iv) the sequence $0 \to A \xrightarrow{\alpha} B \xrightarrow{\beta} C \to 0$ is exact if and only if α is injective and β is surjective and $\operatorname{Ker} \beta = \operatorname{Im} \alpha$. This also means that C and $B/\operatorname{Im} \alpha$ are isomorphic. Since α is injective we can think of it as an inclusion of $A \hookrightarrow B$ of A in B, and then we can identify C with B/A.

Properties iii) and iv) as we will now see are the crucial point for computing homotopy groups.

The way in which exact sequences help us to evaluate homotopy groups is through a result of algebraic topology that assure the existence of a long exact sequence between homotopy groups derived from a short sequence called a *fiber bundle* $F \to E \xrightarrow{p} B$, where p is a projection map.

For what concerns us here the fiber bundle that we consider is very simple and is in fact the special case in which E is a group, let us call it G, F is a subgroup H and B is the coset G/H. Then property iv) above shows us that indeed $H \to G \to G/H$ is an exact sequence.

In this case then one can infer the existence of the long exact sequence

$$\cdots \to \pi_n(H) \to \pi_n(G) \to \pi_n(G/H) \to \pi_{n-1}(H) \to \pi_{n-1}(G) \cdots \to \pi_0(G) \to 0, \tag{6}$$

Then because of property iii) if we know that, for example, $\pi_n(G)$ and $\pi_{n-1}(G)$ are trivial, then we can immediately say that $\pi_n(G/H)$ and $\pi_{n-1}(H)$ are isomorphic.

A nice example of how powerful this approach is, is given by the so called Hopf bundle: $S^1 \to S^3 \to S^2$, where we can think of S^3 as the group of quaternions of unit norm, and S^1 is of course the subgroup of complex numbers with unitary module; then $S^2 = S^3/S^1$. Then the sequence in Eq. (6) becomes

$$\cdots \to \pi_n(S^1) \to \pi_n(S^3) \to \pi_n(S^2) \to \pi_{n-1}(S^1) \to \pi_{n-1}(S^3) \to \cdots, \tag{7}$$

then the above reasoning shows that the isomorphism $\pi_2(S^2) \approx \pi_1(S^1)$ is established once we know that $\pi_2(S^3)$ and $\pi_1(S^3)$ are zero.

As one can imagine the subject, as we have already remarked, is very extended, and what we have discussed here is but a little taste of the whole story. For further references and readings see [29] chapter 2 and 4; Eq. (6) in particular is a special case of Theorem 4.41 proved at page 376.

Degree of a Map. Now to conclude these mathematical preliminaries we will briefly talk about another tool to investigate features of a topological space. Like homotopy it is a topological invariant and classifies maps from two topological spaces.

In the following discussion we will assume some familiarity with the theory of integration over smooth manifolds, since it has been spread in full rigor in the community of physicist ever since General Relativity motivated the detailed study of differential geometry. Interesting discussions can be found in the classic book *Gravitation*, see [39], or for a more mathematically inclined discussion in *Introduction to Smooth Manifolds* (see [34]). A good reference on the subject, that does not make explicit use of concepts of differential geometry can be found in the last chapter of the book by Weinberg *The Quantum Theory of Fields* volume 2 (see [54]).

Now consider a smooth map $F\colon M \to N$ between two orientable compact connected smooth manifolds M and N with the same dimension n.

Then a very simple homotopy invariant of F can be built. Simply take an n-form on N, $\omega \in \Omega^n(N)$, normalized such that $\int_N \omega = 1$. Then such invariant is the integer number $\deg F$ defined by the integral

$$\deg F = \int_M F^*\omega, \tag{8}$$

where $F^*\omega$ is the pullback of ω through F, and is of course itself an n-form: $F^*\omega \in \Omega^n(M)$. The proof that $\deg F$ is an integer is non trivial, and in fact the best way to prove this is to show that the definition we have given here is equivalent to another in which $\deg F$ is characterized as the sum of pluses or minuses (the sign of the Jacobian of the map at some special points in M).

Once the fact that the degree of a map is an integer is established it is also obvious that it is invariant under homotopy, since an integer cannot change under a continuous transformation.

An example to make things more clear is in order. Suppose that F is a map between S^1 and itself, such maps can be classified, as we have shown, on the base of the homotopy class to which they belong. In this case a representative of the k-th class has been identified as the loop that winds around S^1 k times at constant speed. We will now show that the degree of such representative is k.

A normalized 1-form on S^1 in terms of the local coordinate θ (the angular coordinate) is just $\omega = d\theta/(2\pi)$, and the pullback of ω through F is just $F^*\omega = F'(\theta)\, d\theta/(2\pi)$, and since we have assumed that F winds k times around S^1 we have

$$\deg F = \int_0^{2\pi} F^*\omega = \frac{1}{2\pi} F(\theta)|_0^{2\pi} = k. \tag{9}$$

Now we are going to show how all this mathematical tools are used to study topological solitons.

I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed.

— JOHN SCOTT RUSSELL, (1845)

1.2. TOPOLOGICAL SOLITONS

ALTHOUGH the historical roots of soliton physics can be traced back to the study of waves, especially in connection with fluidodynamics (see the above citation), we can probably say that one of the areas in which it has lead to some of the most interesting discoveries is in fact Theoretical Physics (when their relevance has been properly understood).

Between the sixties and the seventies, people discovered — by a careful study of nonlinear equations of motion of some field theories — the existence of soliton-like solutions. One of many interesting features of such solutions is that they are stabilized by the topology of the target space of the fields; for this reason this kind of solutions are usually called topological solitons. It was soon realized that this kind of solutions are particle-like, in the sense that they have all the main properties that are expected from a particle: one can always find a reference frame (by Lorentz boost) in which the field configuration has a finite energy with an energy density that is localized in space: A "lump" of energy. Typically in QFT such particles arise from the quantization of the theory, that is to say they are generated by the introduction of Planck's constant $\hbar$. In soliton-like solutions quantization is mainly due to the topological stabilization that prevents the field configuration to spread out to a constant.

Here we do not want to give a systematic treatment of topological solitons, that is to say we do not want to give an organized classification of solitons and then proceed to a detailed study of them; such analysis can be found for example in the book by Manton and Sutcliffe (see [37]). What we want to do is to give a motivation for what is studied in Chapter 2 and Chapter 3, and to do this we will use a physical model that involves solitons and whose study will make the $\mathbf{C}\mathrm{P}^{N-1}$ naturally emerge. The model in question has been found and discussed in the literature in connection with the problem of confinement in QCD.

The solitons that we will need to discuss in this model are essentially non-abelian vortices and monopoles. So the plan for this section is to briefly introduce the idea of soliton through the simplest known example, the kink solution, and then to make a review of abelian and nonabelian vortices, and of abelian monopoles so that at the end we will be able to understand the model that leads to the $\mathbf{C}\mathrm{P}^{N-1}$ model on a finite worldstrip.

The literature on the subject is vast, especially because this developments, as we have already noticed, are thought to be of primary importance in the matter of confinement, so we cannot hope to cover all the subjects, all the implications of the results and all the conjectures that have been made (especially all those facts that have been established by means of supersymmetric models since they would require a whole per se treatment that is beyond the scope of this work). We hope however to convince the reader that the study of the complex projective model is well motivated by physical reasons.

1.2.1. Kinks

The fist model that every books on topological solitons discusses is the model that has kinks as solutions [37]. This of course is not by chance and it is not due to the historical development of the subject either, it is because this really is probably the simplest model and it is a well equipped playground that will prepare us for the study of more sophisticated theories. So let's have a look at this interesting model.

Consider the theory described by the Lagrangian

$$\mathscr{L} = -\frac{1}{2}(\nabla\varphi)^2 + \frac{1}{2}\dot{\varphi}^2 - U(\varphi), \tag{1}$$

here $\varphi = \varphi(x,t)$ with $x, t \in \mathbf{R}$ (then ∇ here is just a space derivative) and U is a real nonnegative function of φ. It is well known that the Euler-Lagrange equation that can be derived from this Lagrangian through of stationary action principle is the non-linear equation

$$\Box\,\varphi + U'(\varphi) = 0, \tag{2}$$

where $\Box \equiv (\partial_t)^2 - \triangle$. Of course by adding a constant to the Lagrangian — operation that does not change the equation of motion (2) — one can always assume that the global minimum of U is exactly zero.

The first ingredient that is necessary in order to have solitonic solutions, here but also in general, is that the vacuum structure of the model is nontrivial.

Let us suppose that the solutions are constant in time and let us call Ω the set of field configurations that minimize the potential U, that is to say the set:

$$\Omega = \{\, \psi \mid \psi = \arg\min U(\varphi) \,\}, \tag{3}$$

suppose furthermore that Ω is discrete. The assertion that the vacuum structure must not be trivial here is the fact that $\pi_0(\Omega) \neq 0$, so Ω must be made up of more than one point, because otherwise any nontrivial solution could be deformed to the one that is constant everywhere.

Then it is clear that any finite-energy solution must approach an element of Ω as $x \to \infty$ because otherwise the integral of the term $U(\varphi)$ would be divergent. Now two things can happen: Either the values that the fields assume at plus and minus infinity are the same — in that case it is always possible to deform continuously the solution to the one that is constant everywhere at the value prescribed at infinity — or those values are different; this is the case of interest

to us and solutions to the equation of motion that have this characteristic are called kinks.

Let us call $\varphi^{\pm} = \lim_{x\to\pm\infty}\varphi(x)$, and let us suppose that $\varphi^{+} \neq \varphi^{-}$: In this case it is clear that you cannot have a constant energy solution since the term $\varphi'(x)$ cannot be constant, this because of the particular boundary conditions that cannot be modified by continuous transformations.

Another general feature of this system is that one can find solutions of the equation of motion solving a first order differential equation rather than the Euler Lagrange equation (2). This can be done thanks to the topological conditions that kinks undergo. Let us see how to do this here in detail.

The starting point is always a bound of the energy called *Bogomolny bound* [8]. These inequalities, which as we will see can be used essentially every time we study solitons, are derived from a trivial equation of the form $(\mathcal{A}(\varphi))^2 \geq 0$ where $\mathcal{A}(\varphi)$ is an expression that when squared reproduces the energy density plus some other boundary terms. In our case we can take $\mathcal{A}(\varphi) = \varphi'(x)/\sqrt{2} \pm \sqrt{U(\varphi)}$ to obtain the inequality

$$\int_{-\infty}^{\infty}\left(\frac{1}{2}\varphi'^{2}(x) + U(\varphi)\right)dx \geq \pm\int_{-\infty}^{\infty}\sqrt{2U(\varphi)}\cdot\varphi'(x)\,dx, \tag{4}$$

and once we recognize that the term at the left hand side is the energy of the system by changing the variables of integration in the right hand side to $s = \varphi(x)$ we obtain

$$E \geq \pm\int_{\varphi^{-}}^{\varphi^{+}}\sqrt{2U(s)}\,ds. \tag{5}$$

The equality in this expression is obtained when $\mathcal{A}(\varphi) = 0$ that is to say

$$\varphi'(x) = \pm\sqrt{2U(\varphi)}; \tag{6}$$

and this is what is commonly referred to as Bogomolny equation. The solutions to the equation with the plus sign are called kinks, while those corresponding to the minus sign are called antikinks. The proof of the fact that the solutions to Eq. (6) are also solutions of Eq. (2) is quite easy and it can be derived by direct differentiation of this equation.

Till now we have carried on the discussion with a generic potential $U(\varphi)$ that has to satisfy some conditions, however, in order to actually be able to compute some explicit solutions we now consider a specific potential. One can show that

$$U(\varphi) = \lambda(m^2 - \varphi^2)^2 \tag{7}$$

satisfies all the hypothesis we have made on $U(\varphi)$ in the discussion that we have just done. And it turns out to be the simplest potential that gives rise to kinks since the set Ω consists only of two elements: $\Omega = \{-m, m\}$. In this case Eq. (6) becomes

$$\varphi'(x) = \pm\sqrt{2\lambda}(m^2 - \varphi^2) \tag{8}$$

that can be solved directly by separation of variables and leads (for the kink) to

$$\varphi(x) = m\,\mathrm{Th}\left(\sqrt{2\lambda}m(x-a)\right), \tag{9}$$

which has an energy $E = (4/3)m^3\sqrt{2\lambda}$ that is also the rest mass M of the kink. The integration constant a spans what is called in soliton theory the moduli space of the solution; the kink is a solution of a theory which has an invariance under translation: Then we would expect that this invariance is reflected in the solution.

In order to obtain the time dependence of the solution compatible with Eq. (2) we perform a Lorentz boost $\varphi(\gamma(x - vt))$, where v is the velocity of the kink. For non relativistic velocities ($\gamma \approx 1$) the time dependent solution is just a motion on the moduli space. This is in fact a general feature: Under adiabatic perturbations solitons dynamics can be approximated by a motion on the moduli space.

Moduli space motion in this case can be formally derived making the adiabatic ansatz that the time dependence enters only trough the parameter $a(t)$; then direct substitution in the Lagrangian density (1) gives an effective Lagrangian

$$L = \int_{-\infty}^{\infty} \mathscr{L}(x)\, dx = -M + \frac{1}{2} M \dot{a}^2, \tag{10}$$

which yields the equation of motion $\ddot{a} = 0$, that, in turn, implies $a(t) = C + vt$ as we expected.

Now some general remarks on topology. The field φ induces a map φ_∞ between two points (that can be regarded as a zero dimensional sphere S^0) and the vacuum set Ω. Then the various mappings φ_∞ are classified by $\pi_0(\Omega)$, that we supposed to be nontrivial. The degree of such maps defined by Eq. 1.1.2–(8) in this case is just the sum with sign of φ_+ and φ_- with the correct normalization:

$$\deg \varphi_\infty \equiv k = \frac{1}{2m} \int_{-\infty}^{\infty} \varphi'(x)\, dx = \frac{\varphi_+ - \varphi_-}{2m}, \tag{11}$$

and it is sufficient to classify all the solutions since it is 1 for kinks, -1 for antikinks and 0 for constant solutions.

1.2.2. Vortices

We will now talk about one of the two main ingredients that we will need for our model: Vortices. (The word "string" is also commonly used when referring to vortices for reasons we will explain in a moment.)

When we where studying kinks we considered a field theory in one spacial dimension, however such solutions can be extended to any number of dimensions simply by requiring the solution Eq. (9) to be constant along the extra dimensions; for example in two space dimensions one would have domain walls. From this consideration follows that the dimension of the theory is not a relevant parameter to characterize a soliton. What is relevant is the codimension. Kinks are solitons that live in spaces of codimension 1, vortices are characterized by codimension 2, so we will study them in a field theory with dimension two, where they will appear as particle-like solutions. The name strings derive from the fact that if we study those solutions in three spacial dimensions they look like filaments.

Kinks originate in a field theory with translational invariance that is spontaneously broken to a $\mathbf{Z}_2$ symmetry of the vacuum. Vortices, instead, are originated by a theory with rotational invariance with a potential with a circular valley of minima.

1.2.2.1. ANO vortices. Before studying the most known kind of vortices, the Abrikosov-Nielsen-Olesen vortices (see [43]) we will consider a simpler model from which the ANO vortices can be built by a gauging process. This simpler vortex, sometimes called global vortex has, however, the unappealing quality of having an infinite total energy.

The Lagrangian density that gives rise to such solutions is

$$\mathscr{L} = |\partial_\mu \varphi|^2 - U(\varphi), \tag{1}$$

where now $\varphi(x,t)$ is a complex scalar field with $x \in \mathbf{R}^2$ and $t \in \mathbf{R}$. The theory is invariant under rotations $U(1)$ in the complex plane. In order to fulfill the second requirement on the potential that we have mentioned above (the circular valley of minima), we can choose

$$U(\varphi) = \lambda(|\varphi|^2 - v^2)^2. \tag{2}$$

The vacuum manifold of this potential is in fact $\Omega = \{\, \varphi \in \mathbf{C} \mid |\varphi| = v \,\}$ so that while the module of the vacuum solution is fixed its phase is not because $\varphi \to \exp(i\alpha)\varphi$ is a symmetry of the vacuum.

As before we consider static solutions, therefore if there is any hope to have finite energy we must have that at infinity the field is in the minimum of U, that is to say

$$\varphi(r,\alpha) \to v e^{in\alpha} \quad \text{if } r \to \infty, \tag{3}$$

where α and r are the polar coordinates of the spacial plane. This behavior at large r also stabilizes the solution, since it means that the field φ induces a map from $S^1 \to S^1$, that is classified by $\pi_1(S^1) = \mathbf{Z}$. In particular the behavior in Eq. (3) means that φ belongs to the n-th homotopy class. This means, as it is probably intuitive if one thinks about it, that the fields configurations that at infinity make the potential energy vanish are those that wrap around the valley of minima of the potential.

However the behavior in Eq. (3) together with the specific form of the Lagrangian (1) inevitably lead to a divergent energy. In fact for large r we have

$$\partial_i \varphi \sim in\partial_i \alpha = -in\varepsilon_{ij}\frac{x_j}{r^2}, \quad i,j = 1,2 \tag{4}$$

where ε_{ij} is the completely antisymmetric symbol with $\varepsilon_{12} = 1$. Then the integral $\int_{\mathbf{R}^2} (\partial_i \varphi)^2 \, d^2x$ is logarithmically divergent. Such solutions are then physically acceptable only if the domain of the fields is finite (for example a piece of conductor).

In order to have stable, finite energy vortices the missing ingredient are gauge fields. Then the only adjustment that we have to do with the model that we have introduced is to gauge the $U(1)$ symmetry and make it a local symmetry of

the Lagrangian. This can be done of course with the usual procedure of minimal coupling. Introduce a gauge field A_μ with a kinetic term $-1/(4e^2)F_{\mu\nu}^2$ and couple this field to φ by means of the covariant derivative $D_\mu = \partial_\mu - iqA_\mu$, where q is the charge of the field φ in units of e. The Lagrangian of the theory therefore becomes

$$\mathscr{L} = -\frac{1}{4e^2}F_{\mu\nu}^2 + |D_\mu\varphi|^2 - U(\varphi), \tag{5}$$

that is, as we wanted, invariant under local $U(1)$ transformations

$$\varphi \to e^{i\theta(x)}\varphi, \quad A_\mu \to A_\mu + \frac{1}{q}\partial_\mu\theta(x). \tag{6}$$

The potential $U(\varphi)$ must be chosen in such a way that the theory undergoes the Higgs mechanism, we choose for example the form that is written in Eq. (2). In the vacuum usually one chooses the unitary gauge, in which $A_\mu = 0$ and $\varphi = v$, with v real.

In order to have a finite energy one must have that at infinity $|\varphi| \to v$, this in fact ensures that $U(\varphi) \to 0$ as $|x| \to \infty$. The phase of the φ field at infinity is not fixed by this condition, however the details of the phase are unimportant and on topological grounds the only possible asymptotic behaviors of φ are

$$\varphi \to ve^{in\alpha}, \tag{7}$$

where n is an integer. Also the behavior of A_i is fixed by the requirement that $\int_{\mathbf{R}^2} |D_i\varphi|^2 \, d^2x$ must be finite. Then one sees that he must correspondingly transform the field A to get

$$A_i \to \frac{n}{q}\partial_i\alpha = -\frac{n}{q}\varepsilon_{ij}\frac{x_j}{r^2}, \quad i,j = 1,2. \tag{8}$$

(For detailed discussion on the precise derivation of this asymptotic behavior see [37] section 7.3.)

The winding number n here can of course be consistently be calculated using the definition of degree of a map that we gave earlier in Eq. 1.1.2–(8), however because we have that $n/q\partial_i\alpha = A_i$ at infinity, we also have that the winding number coincides with what is called the first Chern number

$$c_1 = \frac{1}{2\pi}\int_{\mathbf{R}^2} B \, d^2x, \tag{9}$$

where B is the magnetic field $B = (1/2)\varepsilon^{ij}F_{ij} = F_{12}$.

Now that we have discussed the topological stability of the vortices we can try to find out the actual solutions. In general the task is not easy because we cannot use the Bogomolny trick and consequently we have to solve the second order equations of motion that are not so easy to handle. However there is a special case in which we can use Bogomolny's first order equations, and that is the case which we will focus on.

We know that due to the Higgs mechanism the vector field A will acquire a mass, that we may call m_V, and we also know that an Higgs field will appear (that is what is left of the φ field) that will have a mass m_H. In general this

two masses will be different, in fact we know how to compute their values and we find that $m_V = \sqrt{2}eqv$ and $m_H = 2\sqrt{\lambda}v$. The limit in which we can perform the Bogomolny trick is when $m_V = m_H$, and the resulting vortex is called BPS vortex (from Bogomolny, Prasad and Sommerfield). In what follows we will put ourselves in this limit.

Then the tension of the string (or the mass of the vortex) becomes

$$T = \int_{\mathbf{R}^2} \left[-\frac{1}{4e^2} F_{ij}^2 + |D_i\varphi|^2 + \frac{q^2 n^2}{2}(|\varphi|^2 - v^2) \right] d^2x \tag{10}$$

Bogomolny trick follows from the observation that the tension density can be rewritten as the sum of two positive terms plus a boundary term; integration of the boundary term can be done immediately using the fact that $c_1 = n/q$ and the expression (9), thus obtaining, after some calculations

$$T = 2\pi n v^2 + \int_{\mathbf{R}^2} \left\{ \frac{1}{2}(B + qe(|\varphi|^2 - v^2))^2 + |(D_1 + iD_2)\varphi|^2) \right\} d^2x. \tag{11}$$

The Bogomolny inequality here reads $T \geq 2\pi n v^2$ and Bogomolny equations are

$$B + qe(|\varphi|^2 - v^2) = 0, \qquad (D_1 + iD_2)\varphi = 0. \tag{12}$$

This two equations can be solved numerically. In the case $n = 1$ due to the symmetry of the problem we can reduce this set of partial differential equations to a system of *ordinary* differential equations making the following guess for the solutions:

$$\varphi(x) = v\phi(r)e^{i\alpha}, \qquad A_i(x) = -\frac{1}{q}\varepsilon_{ij}\frac{x_j}{r^2}(1 - f(r)), \tag{13}$$

where $\phi(r)$ and $f(r)$ are profile functions that because of the boundary behaviors in Eq. (7) and (8) and the requirement of regularity at $r = 0$ they must be such that

$$\phi(\infty) = f(0) = 1 \quad \text{and} \quad \phi(0) = f(\infty) = 1. \tag{14}$$

Substituting Eq. (13) into Eq. (12) one obtains the ordinary differential equations

$$-\frac{1}{\rho}\frac{df}{d\rho} + \phi^2 - 1 = 0, \quad \rho\frac{d\phi}{d\rho} - f\phi = 0, \tag{15}$$

where ρ is the dimensionless quantity $\rho = qevr$. See for example [48] for numerical solutions of the profile functions.

Since ANO strings break translational invariance on the plane, this invariance is restored by the fact that the solutions of Eq. (15) will contain two integration constants that can be thought as the the position of the vortex on the plane.

1.2.2.2. Nonabelian vortices. We now discuss the kind of solitonic solitions that will be central for our future discussion in Section 1.2.4.

The idea here is that while ANO vortices where originated by a gauge theory based on a gauge group that was abelian, namely $U(1)$, nonabelian vortices are originated by gauge theories with a nonabelian gauge group for example $SU(2)$.

What physically really distinguishes such theories from the one we have just studied is the fact that they posses additional moduli that correspond to the possibility of rotations of the magnetic field whose flux form the string: Only solutions with this property will be called nonabelian strings. We make this specification because not all string-like solutions that come from this kind of theories are nonabelian, some of them are just ANO vortices which are stabilized just by a subgroup of the larger gauge group.

Let us now see all this in details in a Yang-Mills theory, with gauge group $SU(2)$. This kind of solutions has been worked out by Auzzi, Bolognesi, Evslin, Konishi and Yung in [5], also while that article was in preparation similar vortices were studied by Hanany and Tong in [28].

In the article [5] it is discussed in details how the model that produces such nonabelian vortices can be derived as an effective field theory (some comments on what an effective field theory is will be made at the beginning of Chapter 2) of $\mathcal{N}=2$ SCQD. Such an effective field theory is shown to be a good approximation of the more fundamental theory at energies that are of order $\sqrt{\mu m}$ where m is of quarks-mass order, while μ is the coupling of a mass term added to the SQCD Lagrangian.

The Lagrangian of such a model is

$$\mathscr{L} = -\frac{F^{a\ 2}_{\mu\nu}}{4g_1^2} - \frac{F^2_{\mu\nu}}{4g_2^2} + |D_\mu \varphi^A|^2 + \frac{g_2^2}{2}\left(\varphi^*_A \frac{\tau^a}{2}\varphi^A\right)^2 + \frac{g_1^2}{8}(|\varphi^A|^2 - 2v^2)^2, \qquad (1)$$

where the fields φ^A for each $A=1,2$ are doublets that transform according to the fundamental representation of $SU(2)$, and $F^a_{\mu\nu}$ is the kinetic term of the gauge fields A^a_μ introduced to make this $SU(2)$ transformation a local symmetry of the theory, in the same way as $F_{\mu\nu}$ is there to take care of the $U(1)$ symmetry. The covariant derivative here is defined by

$$D_\mu = \partial_\mu - \frac{i}{2}A_\mu - \frac{i}{2}A^a_\mu \tau^a, \qquad (2)$$

by comparison with the expression for the ANO covariant derivative we see that here the $U(1)$ charge is $q=1/2$. Besides the local gauge symmetry, if we assume that the fields φ^A mixes between themselves with an $SU(2)$ rotation, the Lagrangian Eq. (1) is also invariant under $SU(2)$ *global* transformations.

Notice also that the couplings to the φ self interactions are functions of g_1 and g_2, and the particular combination that appears in Eq. (1) has been chosen so that the Bogomolny trick can be applied.

The first thing that we have to do is to check out the symmetries that we have pointed out. This discussion is not a mere calculation that can be left to the reader but it somehow belongs to the main logical development since it will equip us with some formalism that will make the analysis of the vacuum easier.

From now on we will refer to the index A as the flavor index, and to the other index (the one that is mixed by the gauge $SU(2)$) as the color index. It is also convenient to collect all the field components into a 2×2 matrix Φ, defined

by

$$(\Phi)_{kA} \equiv \varphi^{kA}. \tag{3}$$

In this notation, color and flavor transformations read respectively

$$\Phi \to U_c \Phi, \qquad \Phi \to \Phi U_f, \tag{4}$$

where $U_c \in SU(2)_c$ and $U_f \in SU(2)_f$; the matter part of the Lagrangian $\mathscr{L}_M$, where $\mathscr{L} = \mathscr{L}_A + \mathscr{L}_M$ reads in this new formalism as

$$\mathscr{L}_M = \mathrm{Tr}[(D_\mu \Phi)^* D_\mu \Phi] - U(\Phi^*, \Phi), \tag{5}$$

where

$$U(\Phi^*, \Phi) = \frac{g_2^2}{2} \mathrm{Tr}\Big(\Phi^* \frac{\tau^a}{2} \Phi\Big) \mathrm{Tr}\Big(\Phi^* \frac{\tau^a}{2} \Phi\Big) + \frac{g_1^2}{8} (\mathrm{Tr}\, \Phi^* \Phi - 2v^2)^2. \tag{6}$$

Now the $SU(2)_c$ symmetry is trivial to check because all we need is the cyclic property of the trace. The flavor symmetry is not so immediate to check since in the first term of U you cannot simply commute U_f with τ^a; however this invariance is instantly evident if we rewrite such term using the following identity:

$$\mathrm{Tr}\Big(\Phi^* \frac{\tau^a}{2} \Phi\Big) \mathrm{Tr}\Big(\Phi^* \frac{\tau^a}{2} \Phi\Big) = -\frac{1}{4} (\mathrm{Tr}\, \Phi^* \Phi)^2 + \frac{1}{2} \mathrm{Tr}[(\Phi^* \Phi)^2]. \tag{7}$$

As we announced, the expression in Eq. (6) is easier to use in order to find the vacuum structure. In fact, as one can see, if we want to make the second term vanish then we should choose Φ proportional to v, while since the Pauli matrices are traceless the vanishing of the first term is achieved by taking the matrix structure of Φ to be the identity. As vacuum configurations one can take

$$\Phi = vI, \quad A_\mu^a = 0, \tag{8}$$

where I is the 2×2 identity; any other configuration that is obtained by a gauge transformation of these fields is just as good as this one.

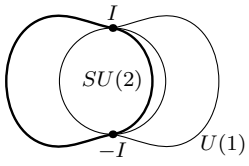
Now we immediately realize that the the vacuum configuration (8) is invariant under a combined color-flavor global transformation

$$\Phi \to U^* \Phi U; \tag{9}$$

this is the key feature that, as we will see, will enable the existence of an additional moduli space of internal orientations. In the literature this property is called *color-flavor locking*.

Let us now try to derive solitonic solutions of this model. First of all notice that, as we have announced, the model supports ANO strings, in fact one has just to ignore the $SU(2)$ gauge group and consider solutions that winds around $U(1)$ at infinity: Those solutions are classified by the elements of $\pi_1(U(1))$. For example for winding number 1 the asymptotic conditions $\Phi(x) \to vIe^{i\alpha(x)}$ and $A_i(x) \to -2\varepsilon_{ij} x_j / r^2$ as $|x| \to \infty$ individuate an ANO string that winds around $U(1)$ once; the factor 2 in the behavior of A is due to the fact that the $U(1)$ charge of φ is $1/2$.

We are not interested in this kind of solutions, although it is somewhat reassuring that they are there, in fact we are looking for solitons that also involve the nonabelian gauge fields A^a_μ. The other part of the gauge group alone, namely $SU(2)$, is not sufficient to stabilize solutions since it is topologically a sphere (i.e. $\pi_1(SU(2)) = 0$). There is however a clever way to achieve topological stability using the whole group; the key observation here is the fact that the center of $SU(2)$ belongs to $U(1)$. More precisely if we call $Z(G)$ the center of the group G — the set of the elements of G that commute with all the other elements in G — one can easily show that $Z(SU(2)) = \{I, -I\} = \mathbf{Z}_2$, and in general $Z(SU(N)) = \mathbf{Z}_N$. Then one can stabilize the solution making it winding at infinity, half on $U(1)$ and the other half on $SU(2)$ as it is pictured in the figure above.



The fact that stability can be achieved in this way from the point of view of the fundamental group, is expressed by the fact that $\pi_1(SU(2) \times U(1)/\mathbf{Z}_2) = \mathbf{Z}$. Notice that this kind of constructions and arguments can be readily generalized to the case of $SU(N)$ since in that case too $\pi_1(SU(N) \times U(1)/\mathbf{Z}_n) = \pi_1(U(N)) = \mathbf{Z}$.

Let us now see how this works in practice for solutions with minimal winding ($n = 1$). Consider the following asymptotic behavior of the Φ field

$$\Phi(x) = \exp\Big(i\frac{1 \pm \tau^3}{2}\alpha(x)\Big) \cdot vI, \qquad |x| \to \infty. \tag{10}$$

Of course this is a vacuum solution since it is obtained from Eq. (8) through a gauge transformation. The factor $\exp(i/2\alpha)$ winds the solution around $U(1)$ while the factor $\exp(\pm i/2\tau_3\alpha)$ winds it around $SU(2)$. Notice also that we consider the case with both the the $+$ and the $-$ since they will give degenerate solutions with the same tension; in the literature those solutions are known as $(1, 0)$ strings (the ones with the $+$) and $(0, 1)$ strings (those with the $-$). Those names comes from the the fact that for unequal masses of the fields φ^A the gauge group becomes $U(1) \times U(1)$ for which the relevant topological classification is given by two integers since $\pi_1(U(1) \times U(1)/\mathbf{Z}_2) = \mathbf{Z}^2$. Correspondingly the asymptotic form of the gauge fields assumes the form

$$A_i(x) = -\varepsilon_{ij}\frac{x_j}{r^2}, \quad A^3_i = \mp\varepsilon_{ij}\frac{x_j}{r^2}, \quad i, j = 1, 2 \text{ and } |x| \to \infty. \tag{11}$$

Because of the peculiar way in which the couplings enter in the Lagrangian (1) we can perform the Bogomolny trick and, exactly as in the case of Eq. 1.2.2.1–(11), we can write

$$T = 2\pi v^2 + \int_{\mathbf{R}^2} \Big\{ \frac{1}{2g_2^2}\Big(F^a_{12} + \frac{g_2^2}{2}\operatorname{Tr}(\Phi^*\tau^a\Phi)\Big)^2 + \frac{1}{2g_1^2}\Big(F_{12} + \frac{g_1^2}{2}\operatorname{Tr}(\Phi^*\Phi - v^2)\Big)^2 + |(D_1 + iD_2)\varphi^A|^2 \Big\} d^2x. \tag{12}$$

Bogomolny inequality then becomes $T \geq 2\pi v^2$, and Bogomolny equations are

$$\begin{cases} \tilde{F}_3^a + g_2^2/2(\varphi^A \tau^a \varphi^A) = 0; \\ \tilde{F}_3 + g_1^2/2(|\varphi^A|^2 - 2v^2) = 0; \\ (D_1 + iD_2)\varphi^A = 0, \end{cases} \tag{13}$$

where we have introduced $\tilde{F}_m = 1/2\varepsilon_{mnk}F_{nk}$ for $n, m, k = 1, 2, 3$ (and in the same way $\tilde{F}_3^a$). In order to find solutions to this equations we restrict ourselves to the case where $A_\mu^1 = A_\mu^2 = 0$ and $\varphi^{kA} \neq 0$ only if $k = A = 1, 2$; and we consider only $(1, 0)$ strings (solutions corresponding to $(0, 1)$ conditions are easily obtainable from them).

At this point we can make the guess

$$\Phi(x) = v \begin{pmatrix} e^{i\alpha}\phi_1(r) & 0 \\ 0 & \phi_2(r) \end{pmatrix} \tag{14}$$

and

$$A_i(x) = -\varepsilon_{ij}\frac{x_j}{r^2}(1 - f(r)), \qquad A_i^3(x) = -\varepsilon_{ij}\frac{x_j}{r^2}(1 - f_3(r)), \tag{15}$$

Eq. (10), and (11) plus regularity impose $\phi_1(\infty) = \phi_2(\infty) = f(0) = f_3(0) = 1$ and $f(\infty) = f_3(\infty) = \phi_1(0) = 0$. Using these guesses Bogomolny equations become the system of ordinary differential equations

$$\begin{aligned} r\frac{d\phi_1}{dr} &= \frac{f + f_3}{2}\phi_1, & r\frac{d\phi_2}{dr} &= \frac{f - f_3}{2}\phi_2; \\ \frac{1}{r}\frac{df}{dr} &= \frac{g_1^2 v^2}{2}(\phi_1^2 + \phi_2^2 - 2), & \frac{1}{r}\frac{df_3}{dr} &= \frac{g_2^2 v^2}{2}(\phi_1^2 + \phi_2^2). \end{aligned} \tag{16}$$

Solutions to those equations can be found numerically (see [5]).

Moduli. Till now we saw that it is possible to find solutions to Bogomolny equations of the form (14) or (15); however as we can directly see from those equations, the vacuum invariance $SU(2)_{c+f}$, the color-flavor locking, is partly broken since the solution retains the invariance only with respect to rotations around the third axis of $SU(2)$ (here by third axis of $SU(2)$ we mean the elements of the group that are proportional to τ^3), in other words the solutions we found implement the breaking pattern $SU(2) \to U(1)$. As it always happens in this situations the additional symmetries that seem to be lost are recovered through the existence of a family of solutions that are described by a nonabelian moduli that parametrize the coset $SU(2)/U(1) \approx \mathbf{C}\mathrm{P}^1$.

Let us see all this in details; the $SU(2)_{c+f}$ invariance fixes the asymptotic behavior of a solution to be

$$\Phi(x) = vI \exp\Big(i\frac{1 + n \cdot \tau}{2}\alpha(x)\Big), \tag{17}$$

where $n \in \mathbf{R}^3$ such that

$$n \cdot \tau = U\tau^3 U^{-1} \tag{18}$$

and $\tau = (\tau^1, \tau^2, \tau^3)$. Moreover Eq. (18) plus the fact that U is unitary implies that $|n| = 1$. The $(1,0)$ and $(0,1)$ strings correspond respectively to $n = (0,0,1)$ and $n = (0,0,-1)$: n is the parametrization of $\mathbf{C}\mathrm{P}^1$.

Now at finite $|x|$ we will have that Eq. (14) changes in the following way

$$\Phi(x) = vU \begin{pmatrix} e^{i\alpha}\phi_1(r) & 0 \\ 0 & \phi_2(r) \end{pmatrix} U^{-1} = v \exp\Big(i\frac{1+n\cdot\tau}{2}\alpha(x)\Big) U \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} U^{-1} \tag{19}$$

with

$$U \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} U^{-1} = U\Big(\phi_1\frac{1+\tau^2}{2} + \phi_2\frac{1-\tau^3}{2}\Big)U^{-1} = \frac{\phi_1+\phi_2}{2} + \frac{\phi_1-\phi_2}{2} n\cdot\tau. \tag{20}$$

Now to conclude our discussion on nonabelian strings we will discuss the worldsheet theory as we did for the kinks.

Worldsheet theory. Similarly to what we have done for the kinks in Eq. 1.2.1–(10), we can ask what the moduli space dynamics is. Here we have four moduli, two that recover the translational invariance of the theory, and two others, that are the ones that we have just discussed above that recover the $SU(2)_{c+f}$ invariance — in what follows we will focus on the orientational moduli described by the vector n.

In order to derive the worldsheet theory we assume that the vector n depends on t and z adiabatically and then one puts the solution we have found with this extra adiabatic dependence into the the original Lagrangian (1). Then the Lagrangian of the effective action can be written down as

$$\mathscr{L} = \frac{\pi}{g_2^2}\big[\dot{n}^2 - (\partial_z n)^2\big], \qquad |n| = 1, \tag{21}$$

that is just the famous $\mathbf{C}\mathrm{P}^1$ or $O(3)$ sigma model of which we will talk about extensively in Section 2.2. The detailed calculations that lead to the explicit derivation of Eq. (21) can be found in the appendix to the third chapter of [48].

Generalization to arbitrary N. In the detailed discussion above we used a $SU(2)$ symmetry, however the argument can be made for a general $SU(N)$ symmetry. The discussion on the symmetries of the model remains the same (except for the fact that in the identity (7) the factor $1/4$ becomes now $1/2N$), and in the vacuum configuration I now is an $N \times N$ matrix. We have already anticipated how topological stability can be reached in this case, and the actual asymptotic behavior of field Φ is similar to the one in Eq. (10) except for numerical constants (the 2 in the denominator becomes an N) and of course the generator τ^3 is replaced by the generator t^{N^2-1}.

Moreover one can prove that the worldsheet Lagrangian here is simply the $\mathbf{C}\mathrm{P}^{N-1}$ Lagrangian that is the main object of study of Chapter 2.

1.2.3. Monopoles

Since we have analyzed in details vortices and nonabelian vortices, and since in the monopole-vortex complex that we will study in the next section much of the

properties of monopoles can be inferred from the vortex, we will only review the basic model from where monopoles, as solitonic solutions emerge — the Georgi-Glashow model — while for a detailed discussion on nonabelian monopoles we'll just refer to [6].

Let us just mention the fact that the idea of magnetic monopoles goes back to the first studies on magnetism when it was observed that a magnets comes always with two poles and it is not possible to isolate those poles. This fact is built right into Maxwell's equation by means of $\nabla \cdot B = 0$.

The matter of the existence of magnetic monopoles and its consistency with electromagnetism and quantum mechanics was addressed by Dirac in 1931 in the article [23], where he formulated the well known Dirac quantization condition. For a concise and clear discussion about Dirac monopoles see [37]. Dirac monopoles, however are not solitons since they have not finite energy; to obtain finite energy monopoles, of solitonic nature, as it was shown by 't Hooft [51] and Polyakov [44], one has to consider Yang-Mills theory coupled to a scalar Higgs field. The resulting soliton looks like the Dirac monopole from far away, however, the solution near the core is such that the singularity is avoided.

1.2.3.1. The Georgi-Glashow model. We will now focus on the $SU(2)$ Yang-Mills-Higgs theory and show how the monopole solution comes out.

The theory is written in terms of a nonabelian gauge field A and an adjoint field ϕ; they both live in the Lie algebra of $SU(2)$. The Lagrangian of the model reads:

$$\mathscr{L} = \frac{1}{4g^2}(F^a_{\mu\nu})^2 + \frac{1}{2}(D_\mu \phi^a)^2 - \lambda(\phi^{a\,2} - v^2)^2, \tag{1}$$

where as usual $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \varepsilon^{abc} A^b_\mu A^c_\nu$; sometimes we may use the matrix notation for the fields ϕ^a: $\phi \equiv \phi^a \tau^a/2$. The BPS limit here is obtained as $\lambda \to 0$.

As we will now discuss, the symmetry breaking pattern of the theory is $SU(2) \to U(1)$. The vacuum configuration for the ϕ field is given by

$$\phi^a = v\delta^{3a}, \tag{2}$$

of course any global color rotation of Eq. (2) is still a vacuum configuration. Then one can say it the other way around: Given a generic vacuum configuration it is always possible to put it into the form written in (2). Notice also that a color rotation around the third axis does not change the the vacuum expectation value of ϕ.

Then we can say that the field ϕ induces a map between the sphere at infinity S^2 to the coset $SU(2)/U(1)$; that is to say into the group $SU(2)$ with two elements identified if they differ only by a rotation $U(1)$ that leaves the ϕ unchanged.

As we have discussed in the mathematical preliminaries, maps from a sphere to a topological space are classified by the second homotopy group. In this case field configurations are classified by $\pi_2(SU(2)/U(1)) = \mathbf{Z}$.

Assuming the BPS limit, the energy (mass) of the monopole can be read off from Eq. (1):

$$E = \int_{\mathbf{R}^3} \left(\frac{1}{2g^2} (B_i^a)^2 + \frac{1}{2} (D_i \phi^a)^2 \right) d^3x, \tag{3}$$

where we have defined $B_i^a \equiv -(1/2)\varepsilon_{ijk} F_{jk}^a$. Since we are considering static field configurations it seems natural to assume $A_0^a = 0$, this assumption will be verified a posteriori.

Since we are in the BPS limit, we can as usual rewrite the energy as a square of something plus a boundary term, in this case we simply have that

$$E = \int_{\mathbf{R}^3} \left[\frac{1}{2} (B_i^a/g - D_i \phi^a)^2 + \frac{1}{g} B_i^a D_i \phi^a \right] d^3x, \tag{4}$$

the term $B_i^a D_i \phi^a$ is actually a boundary term since

$$\frac{1}{g} \int_{\mathbf{R}^3} B_i^a D_i \phi^a \, d^3x = \frac{1}{g} \int_{\mathbf{R}^3} \partial_i (B_i^a \phi^a) \, d^3x = \frac{1}{g} \int_{S^2_\infty} B_i^a \phi^a \, dS^i \equiv q_M, \tag{5}$$

where S^2_∞ is the sphere that surrounds $\mathbf{R}^3$ at infinity, and q_M is the monopole charge since the last integral in Eq. (5) is the flux of the *gauge invariant magnetic field* $\mathcal{B}_i \equiv B_i^a \phi^a / v$. On a sphere far away, the finite energy condition imposes that $\mathcal{B}_i = B_i^3$.

Then the energy becomes

$$E = q_M + \frac{1}{2} \int_{\mathbf{R}^3} (B_i^a/g - D_i \phi^a)^2 \, d^3x, \tag{6}$$

meaning that the Bogomolny bound is $E \geq q_M$, with Bogomolny equations

$$B_i^a = g D_i \phi^a. \tag{7}$$

Now if we want to study the solutions characterized by $n = 1$, we immediately realize that we can just identify the sphere at infinity with the group sphere, so that we can take

$$\phi^a = v \frac{x^a}{r}, \quad |x| \to \infty, \tag{8}$$

moreover, to obtain a finite energy we must also require that $D_i \phi^a$ falls off at infinity sufficiently rapidly; this requirement fixes the asymptotic behavior of A_i^a:

$$A_i^a = \varepsilon^{aij} \frac{x^j}{r^2}, \quad |x| \to \infty. \tag{9}$$

The knowledge of the asymptotic behavior of the fields ϕ and A is sufficient to evaluate the magnetic charge defined in Eq. (5). In fact we have that $\mathcal{B}_i = x^i / r^3$ at large $|x|$, and then inspection of Eq. (5) immediately tells us that $q_M = 4\pi/g$.

To obtain solutions to Bogomolny equations at finite $|x|$ we use the guess

$$\phi^a = v h(r) \frac{x^a}{r}, \qquad A_i^a = \varepsilon^{aij} \frac{x^j}{r^2} f(r), \tag{10}$$

with $h(\infty) = f(\infty) = 1$ and $h(0) = f(0) = 0$. Substituting these expressions for ϕ and A into the Bogomolny equation Eq. (7) we obtain two ordinary *nonlinear* differential equations for h and f:

$$\begin{cases} f'(\rho) = h(\rho)(1 - f(\rho)); \\ h'(\rho) = 1/\rho^2(2f(\rho) - f^2(\rho)), \end{cases} \tag{11}$$

for which we know the analytical solutions

$$f(\rho) = 1 - \frac{\rho}{\mathrm{Sh}(\rho)}, \qquad h(\rho) = \frac{1}{\mathrm{Th}(\rho)} - \frac{1}{\rho}. \tag{12}$$

With this solutions at hand one can check by direct calculations that the charge of the monopole is $q_M = 4\pi/g$.

1.2.4. Nonabelian monopole-vortex complex

In the introduction of the chapter we argued about how we can find systems characterized by a M-V-M configuration made up of a nonabelian vortex and two monopoles. We will now discuss in more details how this is realized.

The basic principle behind this is the existence of a hierarchical symmetry breaking, in which, thanks to the first symmetry breaking pattern, monopoles are generated, while the second breaking gives rise to the vortex.

To be more precise consider a system with the following symmetry breaking pattern:

$$SU(N+1)_c \times SU(N)_f \xrightarrow{v_1} SU(N)_c \times U(1)_c \times SU(N)_f \xrightarrow{v_2} SU(N)_{c+f}, \tag{1}$$

where as usual c and f make distinction between the color (local gauge) and the flavor symmetry, and v_1 and v_2 are the scales at which the symmetry breaking takes place. Intuitively $1/v_1$ and $1/v_2$ are respectively the characteristic sizes of the monopoles and of the vortex: We choose $v_1 \gg v_2$. As we saw the fact that $\pi_2(SU(N+1)/(SU(N) \times U(1))) = \mathbf{Z}$ supports the existence of magnetic monopoles, while the low energy symmetry breaking pattern with $\pi_1(SU(N) \times U(1)/\mathbf{Z}_N) = \mathbf{Z}$ suggests the existence of vortices.

Strictly speaking, neither the monopoles, nor the vortices exist in this theory as configurations that are not stabilized by topology since

$$\pi_2(SU(N+1)) = \pi_1(SU(N+1)) = 0. \tag{2}$$

However since the system is considered at different scales this allows us to consider effective theories at each scale where vortices and monopoles will be stabilized by topology.

Eq. (2) is an indication that the vortex must end (since it must not be detected by loop retraction). Actually the vortex and the monopoles are strictly related by the exact sequence in Eq. 1.1.2–(6) in particular by the segment

$$\pi_2(SU(N+1)) \to \pi_2\Big(\frac{SU(N+1)}{SU(N) \times U(1)}\Big) \to \pi_1(SU(N) \times U(1)) \to \pi_1(SU(N+1)), \tag{3}$$

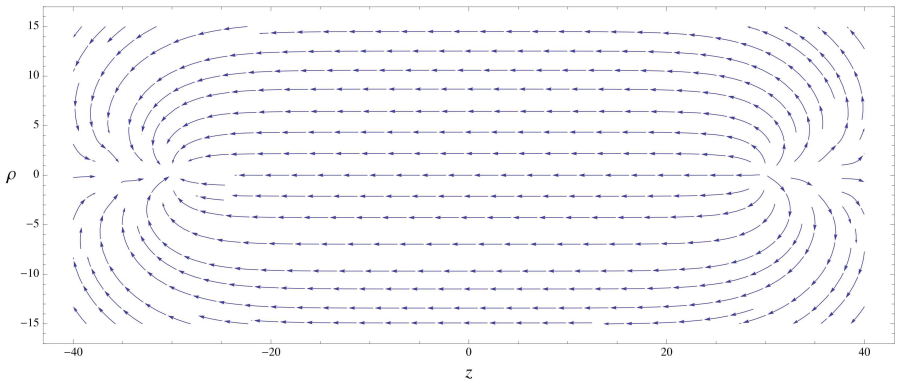


Fig. 1. Numerical simulation of the M-V-M complex in cylindrical coordinates, the image is taken from [17].

then because of Eq. (2) we conclude that $\pi_2(SU(N+1)/(SU(N)\times U(1))) \approx \pi_1(SU(N)\times U(1))$. This means that each minimum vortex that belongs to an equivalence class in $\pi_1(SU(N\times U(1)/\mathbf{Z}_N))$ ends at a monopole of the corresponding equivalence class of $\pi_2(SU(N+1)/(SU(N)\times U(1))))$. Of course this kind of connection is such that the magnetic monopole is equipped with the same $\mathbf{C}\mathrm{P}^{N-1}$ moduli space that is characteristic on the nonabelian vortex as we have seen in Section 1.2.2. One explicit check of this property can be done by verifying that the flux through a small sphere around a monopole matches exactly (is equal to) the flux along the vertex, this check has indeed been done in [7]. Also in [17] starting from a $\mathcal{N}=2$ supersymmetric model, a numerical solution to the exact Bogomolny equations has been found, the result of such numerical analysis is shown in Fig. 1.

More recently Chandrasekhar Chatterjee and Kenichi Konishi studied the monopole-vortex complex in the large distance approximation (see [15]) in the same article they propose an essential model that presents a hierarchical symmetry breaking pattern like the one in Eq. (1). In particular the model they considered is defined by the Lagrangian

$$\mathscr{L} = -\frac{(F_{\mu\nu})^2}{4} + |D_\mu\phi|^2 + |D_\mu\varphi^A|^2 - U(\varphi,\phi), \tag{4}$$

where ϕ is a scalar field that transforms according to the adjoint representation of $SU(N+1)$; the φ^As, with $A=1,\dots N_f \equiv N$, are a set of scalar fields in the fundamental representation. In order to achieve the wanted hierarchical symmetry breaking the crucial ingredient is the form of the potential, that can be taken to be

$$U(\varphi,\phi) = |\mu\phi^B + \varphi^*_A T^B \varphi^A|^2 + |(T^B\phi^B + m_A)_{jk}\varphi^{kA}|^2, \tag{5}$$

where m_A are the masses of φ and the parameter $\mu \ll m_A$.

Now, in order for this potential to vanish, and because of the second term, φ must be vanishing or either an eigenvector of ϕ with eigenvalue m_A. We choose the equal mass limit $m_A = m_0$ for all A and we take the vacuum of ϕ to be

$$\langle\phi\rangle \equiv \langle T^B \phi^B\rangle = m_0 \begin{pmatrix} N & 0 \\ 0 & -I_N \end{pmatrix}. \tag{6}$$

This expectation value breaks the $SU(N+1)$ gauge symmetry down to $SU(N)\times U(1)$, as we wanted; it also make the first color component of the φ^A become massive with mass $v_1 = m_0(N+1)$ and forces

$$\langle\varphi^A\rangle = v_2 I, \qquad v_2 = \sqrt{2(N+1)\mu m_0} \ll v_1. \tag{7}$$

With this, one achieves the low energy (order of v_2) breaking pattern $SU(N)_c \times U(1)_c \times SU(N)_f \to SU(N)_{c+f}$. In the same article, after a detailed analysis of the M-V-M complex, the orientational zeromodes and the worldsheet theory are discussed too.

The key observation here is that although the vacuum of the theory is invariant under $SU(N)_{c+f}$ (it is color-flavor locked), the monopole-vortex solution breaks this symmetry down to $SU(N-1)\times U(1)$. (This it is exactly what happened for vortices since the monopoles follow the same orientation of the vortex.) Then, as usual, the symmetry is recovered because of the existence of a moduli space that parametrizes the set of continuous configurations that span the coset

$$\frac{SU(N)}{SU(N-1)\times U(1)} \approx \mathbf{C}\mathrm{P}^{N-1}. \tag{8}$$

At this point the moduli space can be parametrized using the so called reducing matrix [15]; assuming an adiabatic time dependence, one can show that the effective action we obtain in this way is the one of the well known two dimensional $\mathbf{C}\mathrm{P}^{N-1}$ model defined on a *worldstrip* with the boundaries defined by the worldline of the monopole and antimonopole. Strictly speaking then the total action is the two dimensional $\mathbf{C}\mathrm{P}^{N-1}$ model action (for the vortex) plus the one dimensional $\mathbf{C}\mathrm{P}^{N-1}$ action (for the monopoles).

This is what at the end of the day motivates us to carefully investigate the two dimensional $\mathbf{C}\mathrm{P}^{N-1}$ model on a strip.

CHAPTER TWO

COMPLEX PROJECTIVE MODEL

In fact, everything we know is only some kind of approximation, because we know that we do not know all the laws *as yet. Therefore, things must be learned only to be unlearned again or, more likely, to be corrected.*

— RICHARD P. FEYNMAN, The Feynman Lectures on Physics (1963)

But the ultimate goal is to learn some basic wisdom that we will someday put to use in answering that underlying question: How does Nature work?

— SEAN M. CARROLL (2015)

THE CHIEF PURPOSE of this chapter is to study the complex projective model, the $\mathbf{C}\mathrm{P}^{N-1}$ model and the solution of such model in the large N limit. As we have already seen in Section 1.2.4 this model arises from the study of internal moduli dynamics of nonabelian vorticies and monopoles. However, as it is remarked at the beginning of Chapter 1, curiously enough, such model was already known in connection to Quantum Cromodynamics and has been extensively studied in that context.

One can ask himself why we need to study different models from the ones we believe actually describe Nature; in the last few years there have been many experimental confirmations of the fact that the Standard Model is the correct theory that describes — at least at the energies that are currently available to us — the spectrum of particles that populate our world and, more importantly, how they interact. So why don't we concentrate all our energies into the study of this theory?

There are several good reasons not to do so and instead making efforts to understand all this variety of quantum field theory models.

Probably the most convincing one is the fact that some of these theories are actually *effective* theories for the more "fundamental" ones (that covers a wider range of energies). An effective theory is a model that replaces some other theory that you are studying, pretty much in the same spirit of the effective potential that it is used in classical mechanics to reduce complicated three dimensional central-force problem down to a one dimensional problem.

With this kind of analysis one reduces a possibly complicated theory to a possibly simpler one, that is valid (that is to say gives answers that are compatible with the whole theory) in a certain range of energies.

Within the framework of Lagrangian theories, most of the times, effective theories can be derived writing down the most general theory that is compatible with the symmetries of the model from which we start. Typically effective field theories describe what happens at *low energies* where we do not need all the details in the short range interactions provided by the full fundamental theory. In this regime it is clear that the other criterion one should use to build such theories, other than symmetry, is to keep only terms with the lowest number of derivatives, because higher derivatives terms at low energies are highly suppressed.

Let us make an example. Suppose you want to study pion scattering. Pions are mesons, this means that are made of quarks, and the theory that describe how quarks interact with each other is QCD. In principle one should use that theory, or better the techniques that commonly goes under the name of *current algebra* methods (see [54]). Unfortunately, as one can guess, such calculations are very lengthy and it is difficult to deal with symmetries that are not exact, like chiral symmetry of quantum chromodynamics.

Since we are interested in the pions, such degrees of freedom are described by the effective Lagrangian

$$\mathscr{L} = \frac{1}{2}\frac{\partial_\mu \vec{\pi} \cdot \partial^\mu \vec{\pi}}{(1 + \vec{\pi}^2/F^2)^2}. \tag{1}$$

This is a non-linear sigma model for the symmetry breaking pattern $SU(2) \times SU(2)$ that breaks down to $SU(2)$.

However there is also another reason to keep studying simple models: We need to extend our knowledge of quantum field theory beyond perturbation theory.

One of the reasons quantum field theory is so hard to understand is because it is far from our experience since there are very few simple examples that can be studied in details, and even though the successes of perturbation theory must not be forgotten it is clear that we cannot rely entirely on this approach.

Before going on with the analysis of the complex projective model it is instructive, in order to appreciate best the results that can be obtained in the study of models in the large N limit, and to gain awareness of how subtle is the study of the quantum theory of fields, to revise some basic problems of perturbative expansions. In particular we will try to discuss how Dyson argument against convergence of perturbative series is realized in the simple φ^4 theory in various dimensions.

In this kind of discussion we will follow the presentation given in [49].

Dyson's argument. Dyson addressed the problem of convergence of the *renormalized* perturbative series in quantum electrodynamics. The main idea of his argument is as follows; consider the expansion around $e^2 = 0$ of a physical quantity F:

$$F(e^2) = F(0) + e^2 F_2(0) + e^4 F_4(0) + \ldots, \tag{2}$$

where F_2 and F_4 are proportional to the derivatives of F.

Suppose that this series is convergent, then there must exist a neighborhood of the origin in which F is analytic. This means that it makes sense to evaluate such a quantity for $e^2 < 0$; on the other hand changing the sign of the coupling is equivalent to consider a physics in which particles with opposite sign repels each other and particle with the same charge attracts each others. Therefore $F(-e^2)$ will be the value of the physical quantity that we are considering measured in a world in which the electromagnetic interaction changes nature in the way we have just described. The point is that with the ordinary interaction if one starts to collect together particles with opposite sign, due to stability of matter the energy that such configuration can ever reach is bounded from below.

However in a theory in which the electromagnetic coupling is $-e^2$ you can indefinitely create pairs of particles (by means of an external perturbation) and then lower the energy grouping particles with the same charge into two well separated clusters. This means that the vacuum of the theory is no more stable.

This drastic change of behavior of the theory cannot be reconciled with the analyticity of F in a neighborhood of the origin, and therefore goes against the assumption of the convergence of the renormalized perturbation series.

Just to give a concrete example of how this can happen, consider the theory of a scalar field with quartic self interaction in zero space and zero time dimensions with mass μ. In such a theory the Euclidean functional integral assumes a very simple form

$$Z(\mu, g) = \int_{-\infty}^{\infty} e^{-\mu\varphi^2 - g\varphi^4}\, d\varphi, \tag{3}$$

where μ and g are two real positive parameters. Let us consider the expansion of such quantity in powers of the coupling as usual:

$$Z(\mu, g) = \sum_{n\geq 0} Z_n(\mu) g^n, \tag{4}$$

the coefficients are just

$$\begin{aligned} Z_n(\mu) = \frac{1}{n!}\frac{\partial^n Z}{\partial g^n}(\mu, 0) &= \frac{(-1)^n}{n!}\int_{-\infty}^{\infty} e^{-\mu\varphi^2}\varphi^{4n}\, d\varphi \\ &= \frac{(-1)^n}{n!}\mu^{-2n-1/2}\int_0^{\infty} t^{2n-1/2}e^{-t}\, dt \\ &= \frac{(-1)^n}{n!}\mu^{-2n-1/2}\Gamma(2n+1/2), \end{aligned} \tag{5}$$

where the passage from the first to the second line is achieved by the change of variable $\mu\varphi^2 = t$. Because we also have that $\Gamma(2n+1/2) = 2\sqrt{\pi}\, 2^{-4n}(4n)!/(2n)!$, putting all together we immediately obtain the large n behavior for $Z_n(\mu)$:

$$Z_n(\mu) \sim 2^n n!, \quad \text{for large } n. \tag{6}$$

This tells us that the perturbative series diverges; in fact these coefficients do not satisfy the analyticity condition $|Z_n(\mu)| \leq c^n$, for some constant c.

Actually we can say something more: The perturbative series is asymptotic to the exact solution. This fact that showed up in this simple theory could be

the explanation of why the sum of the lowest terms give a good approximation near the origin.

A series $\sum_n a_n x^n$ is said to be asymptotic at the origin to a function $f(x)$, that is C^∞ in $(0\,..\,\epsilon)$ if for every k, $f(x) - \sum_{n=0}^{k} a_n x^n = o(x^k)$ as $x \to 0^+$. Indeed

$$\left| Z(\mu, g) - \sum_{n=0}^{k} Z_n(\mu) g^n \right| \leq C_k g^{k+1} |Z_{k+1}(\mu)|. \tag{7}$$

Non-perturbative study of φ^4 in $0+1$, $1+1$ and $2+1$ dimensions has been studied, however the analysis (especially for the last two cases) is not trivial, so we won't discuss it here. In the $1+1$ and $2+1$ cases it has been proven that scattering matrix elements have divergent perturbative series and it has been shown that Schwinger functions are Borel summable. For completeness we recall that a series $\sum_n a_n z^n$ is Borel summable if and only if the following properties holds:

i) $|a_n| \leq C^{n+1} n!$
ii) its Borel transform $B(x) \equiv \sum_n a_n x^n / n!$, defined for $|x| < C^{-1}$ has analytic continuation to a neighborhood of $[0\,..\,\infty]$, and for $x > 0$, $|B(x)| \leq e^{Dx}$ for some D,
iii) the integral

$$f(x) = \frac{1}{x} \int_0^\infty e^{-y/x} B(y)$$

converges absolutely.

Recently it has also been proven [49] that φ^4 in $3+1$ dimensions is trivial, so that it seems that the perturbative series is far away from the actual behavior of the whole theory.

We now turn to the study of the large N limit and try to give some evidences that solutions of theories in this limit explicitly incorporate crucial non-perturbative effects. This, together with many other appealing features makes of this kind of expansion introduced by 't Hooft an interesting laboratory for the study of quantum field theory beyond perturbation theory.

Just then her head struck against the roof of the hall: in fact she was now more than nine feet high, and she at once took up the little golden key and hurried off to the garden door.

— LEWISS CARROLL, Alice's Adventures in Wonderland (1865)

Remember, with great power comes great responsibility.

— Uncle Ben, Spider-Man (2002)

2.1. LARGE N LIMIT

As we already partially argued the 't Hooft expansion is a very useful tool to get information about a given theory especially when ordinary perturbation theory cannot be trusted even at the lowest orders. But what exactly is this limit?

As the name might suggest we have to deal typically with a symmetry that mixes N fields between themselves. This expansion also called $1/N$ expansion was originally introduced by Gerardus 't Hooft [50] and later developed by Witten [57] in order to understand some properties of QCD. In such theories (Yang-Mills theories) even though it is not possible to carry on explicit computation, even at the lowest order, this limit enables the possibility to infer general properties; noteworthy are also the properties that one can infer on mesons.

At the same time, the large N limit makes exactly soluble some models that can give us some solid insight on a certain number of phenomena. In our case our interest is directed towards the $\mathbf{C}\mathrm{P}^{N-1}$ model.

In order to present this theory in some details we will introduce the large N expansion through a theory of a vector of scalar fields $\phi = (\phi^1, \phi^2, \dots, \phi^N)$ with a $O(N)$ symmetry and quartic self interactions; this model will only serve to introduce the basics of $1/N$ expansion, and especially the techniques that one can use to count the powers of N.

We will then discuss how one must change this procedures when there is a gauge symmetry with fields in adjoint representation.

Then in the next section we will introduce and carry on a systematic study of the $\mathbf{C}\mathrm{P}^{N-1}$ model.

2.1.1. Expansion basics

As we have just seen we will use a ϕ^4 model to explain the basics of $1/N$ expansion, or if you want the combinatorics of Feynman diagrams. This analysis is taken from the book by Coleman [18].

This theory is specified by the following Action

$$S[\phi] = \int d^D x \left\{ \frac{1}{2}\partial_\mu \phi \cdot \partial^\mu \phi - \frac{m^2}{2}\phi \cdot \phi - \frac{\lambda}{8}(\phi \cdot \phi)^2 \right\} \equiv \int d^D x \, \mathscr{L} \qquad (1)$$

in D dimensions. Suppose that $D \leq 4$, and let us take the components of ϕ, ϕ^a to be real $a = 1, \dots, N$.

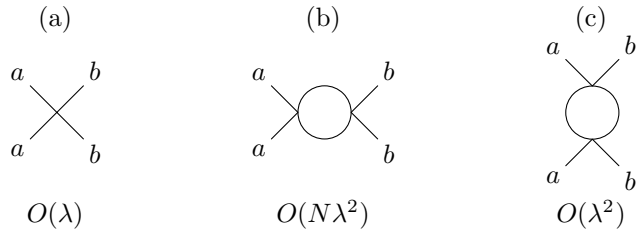
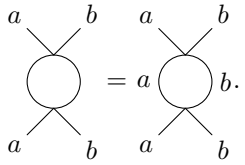


Fig. 1. First relevant graphs for meson-meson scattering up to $O(\lambda^2)$. Letters on loops are omitted, so that in general these are classes of graphs.

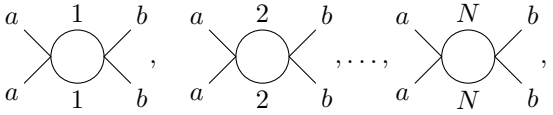
Of course this theory is invariant under $O(N)$ transformations of the fields; that is to say under $\phi \to \phi' = M\phi$ where M is a $N \times N$ dimensional matrix in $O(N)$.

Suppose one wants to consider meson-meson scattering in which two type a mesons are scattered into two type b mesons ($aa \to bb$), then the first contributions in ordinary perturbation theory are those shown in Fig. 1. Notice that the graphs in Fig. 1(a) and Fig. 1(c) are just single diagrams, since the indices carried by the internal lines are uniquely determined by the indices that appear on the external lines; for the graph in Fig. 1(a) this is obvious, but also



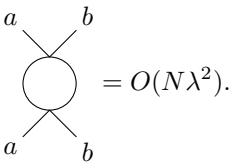
(2)

However, for each index $c = 1, \ldots, N$, one can build a graph like the one in Fig. 1(b), since the index that flows in the loop is independent of the indices carried by the external legs. So actually this graph represents a class of Feynman diagrams:



(3)

and each of this diagrams will give a contribution of order λ^2, so that actually



(4)

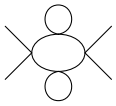
At this point it seems a silly proposal to evaluate diagrams at large N since, of course, all diagrams with more loops cannot be ignored. However everything depends on the question we are asking: suppose that we introduce the new coupling

$$g \equiv \lambda N, \tag{5}$$

the so called 't Hooft coupling, then we can ask to compute diagrams at fixed g and large N. This question as we will see really makes sense.

With this redefinition we have that  $= O(g/N)$,  $= O(g^2/N)$, while  $= O(g^2/N^2)$. This shows that there is not a simple relation between the numbers of loops, or the order in the g coupling, and the order in $1/N$, and it becomes pretty clear that the difficult problem here is to give a characterization of the various orders of the large N expansion.

Let us consider some more examples before we try to prove that the choice we have made of the coupling indeed does not give a trivial expansion or one with terms proportional to positive powers of N. From now on we will not put labels on the arcs of the graphs, because otherwise the notation becomes too cumbersome. For example we have

 $= O(\lambda^4 N^3) = O(g^4 N^{-4} N^3) = O(g^4/N),$ (6)

but also

 $= O(\lambda^4 N^3) = O(g^4/N).$ (7)

This two examples show that there are actually an infinite number of diagrams that are order $1/N$, in fact we can always show an infinite subset of this kind of diagrams: $S = \{ (\text{})^n \mid n = 0, 1, 2, \dots \}$ with the understanding that, as it is showed in Eq. (7), $(\text{})^n$ means the graph with n side by side loops. It is in fact immediate to prove that each element in S is of order $1/N$: the generic element of this set has $k+1$ vertices and k loops with an independent color index on each one, then its order is $\lambda^{k+1} N^k$, that if it is written using the 't Hooft coupling gives exactly g^{k+1}/N. Of course not all diagrams that are $O(1/N)$ belong to S (for example the diagram in Eq. (8)), there are many more!

Now that we have this additional graphical intuition we must go on to devise a general proof, that will turn out to be also the usual way with which we can handle in general the large N limit. The idea is to introduce an auxiliary, non-dynamical field ω that will change Feynman rules of the theory in such a way that it will be easier to count powers of $1/N$.

Consider the following modification of the Lagrangian

$$\mathscr{L} \to \mathscr{L}' = \mathscr{L} + \frac{1}{2}\frac{N}{g}\left(\omega - \frac{g}{2N}\phi \cdot \phi\right)^2, \tag{8}$$

then the claim is that this gives the same theory. In fact notice that such a modified theory does not contain derivative terms $\partial_\mu \omega$, so the Euler-Lagrange equations for ω reduces to

$$\omega - \frac{g}{2N}\phi \cdot \phi = 0; \tag{9}$$

using this equation the additional term in Eq. (8) cancels.

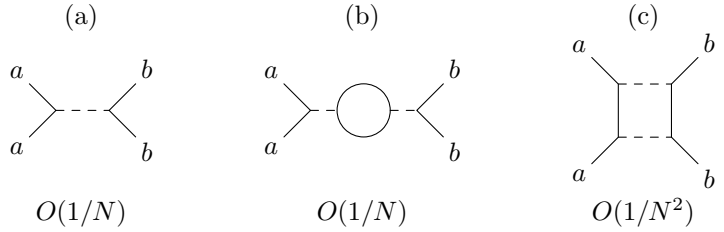


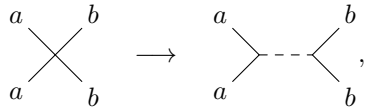
Fig. 2. Graphs in Fig. 1 written using Feynman rules obtained from $S[\phi,\omega]$. The order of the diagrams in the $1/N$ expansion is easily obtained evaluating the difference between the number of propagators of the field ω and the number of ϕ-loops.

So now our theory is described in terms of the new Action

$$S[\phi,\omega]=\int d^Dx\left\{\frac{1}{2}\partial_\mu\phi\cdot\partial^\mu\phi-\frac{m^2}{2}\phi\cdot\phi+\frac{N}{2g}\omega^2-\frac{1}{2}\omega\phi\cdot\phi\right\}, \tag{10}$$

notice that now the factor N appears in the propagator of the field ω (that turns out to be proportional to $1/N$) and, as usual, we get a contribution of a factor N for each ϕ loop.

In this theory the field ω "mediate" the quartic ϕ interaction so that diagrams in this notation are obtained from the previous diagrams using the substitution



(11)

for each vertex in the diagrams. In Fig. 2 this substitution has been done for the diagrams in Fig. 1.

Taking into account the correct position of the indices we can as well rewrite the graphs in Eq. (6) and Eq. (7).

By inspection of Eq. (10), and looking at the few examples we have analyzed so far we realize that we can do even better. We can write an effective theory for the fields ω; this can be understood very well from the point of view of the functional integral. The partition function for the initial theory is

$$Z(g,m^2)=\int[d\phi][d\omega]\,e^{iS[\phi,\omega]}. \tag{12}$$

Since the field ϕ appears only quadratically in the action (10), we can integrate those fields away, what is left can be considered as an effective theory for the field ω. In formula

$$e^{iS_{\rm eff}[\omega]}\equiv\int[d\phi]\,e^{iS[\phi,\omega]}, \tag{13}$$

this calculation involves the computation of a gaussian integral that will produce a factor of the form $(\det(\mathcal{M}))^{-N}$; we won't derive the detailed form of this term right now since we will do a similar calculation while studying the $\mathbf{C}\mathrm{P}^{N-1}$ model. However this observation together with the fact that the only term in $S[\phi,\omega]$ that is not quadratic in ϕ is proportional to N, is sufficient to conclude that

$S_{\text{eff}}[\omega]$ is proportional to N. This property makes the counting of powers of N straightforward.

The diagrams induced by $S_{\text{eff}}[\omega]$ are therefore obtained from the diagrams of $S[\phi, \omega]$ cutting away the external ϕ-legs and by doing all integrals over the ϕ loops (this leaves only non-local interactions between the ω-fields).

Now consider a general graph in the theory defined by $S_{\text{eff}}[\omega]$, it will have E external lines, I internal lines, V vertices and L loops. As it is well known these quantities are not independent but they are related by a "topological" relation that holds regardless of the theory, of how many kind of propagators or interaction vertices there are; the only requirement is the connectivity of the graph. Such relation is

$$L = I - V + 1. \tag{14}$$

Now because of the observation we previously made on the linear dependence of S_{eff} on N, each propagator will carry a factor $1/N$ and this will contribute in the graph with a factor N^{-E-I} and each vertex will contribute with a term proportional to N, and this will add a factor N^V, so the total dependence of a general graph on N is

$$N^{-E-I+V} = N^{1-E-L}, \tag{15}$$

where in this equality we have used the topological relation (14).

Eq. (15) tells us that diagrams that contribute with a positive power of N must have $1-E-L > 0$, since in meson-meson scattering $-E < -2$, we conclude that $L < -1$ is a nonsense. This proves that the large N expansion is actually well defined in the terms we have specified.

Meson-meson scattering then will be given by all diagrams with two external ω legs. We need to be careful only of one detail, because Eq. (10) has a linear term in ω, the effective theory will have a linear vertex, that is to say in general $\delta S_{\text{eff}}/\delta\omega|_{\omega=0} \neq 0$, from a diagrammatic point of view this means that given a diagram of a certain order in $1/N$ we can always construct infinite others of the same order just by adding ω propagators that terminate in the linear vertices. This problem can be however easily solved, as it is well known, just by shifting the value of ω by a constant.

2.1.2. Large N in QCD: Adjoint representation

Let us now discuss how one has to modify the above discussion when one wants to apply the large N limit to a theory with fields that live into the adjoint representation.

We could study the same theory with the ϕ field in the adjoint representation, however it is much more interesting to study how we can apply the large N expansion to QCD, since it reproduces known properties of the hadronic world.

This time the combinatorics of diagrams become more difficult to derive due to the fact that we have to handle more indices. In order to make systematic progresses, we need to introduce a nice mathematical concept that classifies topological spaces that is called Euler characteristic. So now we will make a short detour to introduce this concept using a general mathematical definition.

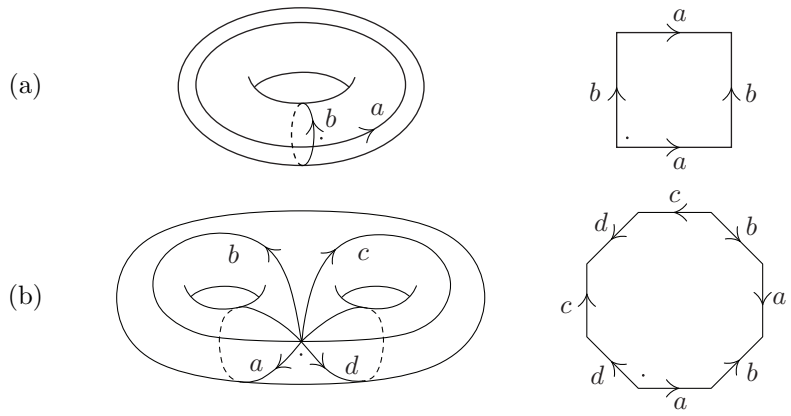


Fig. 1. An orientable surface M_g, of genus g can be constructed gluing appropriately the sides of a $4g$-gono. The dot indicates where that angle is mapped in the gluing process.

We will use cell complexes to define this topological invariant not only because it is a nicer definition that generalizes the usual $vertices - edges + faces$ formula but also because the way in which we will analyze Feynman graphs is exactly the way in which we built spaces from CW complexes.

Cell Complexes. In order to give a sound definition of Euler characteristic we need to introduce the concept of cell complexes, that is a general procedure to construct spaces.

Let us take inspiration from a very nice procedure that can be used to construct orientable surfaces M_g of genus g (remember that the genus of a given surface is equal to the number of handles).

As it is shown in Fig. 1(a) a torus $S^1 \times S^1$ can be obtained by identifying opposite sides of a square. As Fig. 1(b) suggests this procedure can be generalized to an arbitrary orientable surface of genus g just starting from a polygon with $4g$ sides.

This way of building M_g can be equally described as follows: Take a point, that we will call a 0-*cell*, and attach to this point 2 arcs (1-*cells*), so far we have what is called the 1-*skeleton*, then finally attach to the arcs an open disk, or 2-*cell* that is exactly the interior of the polygon, this gives the 2-*skeleton*.

This procedure can be made rather general, and for this reason defining Euler characteristic in terms of CW complexes is convenient.

We will now describe the general inductive procedure to construct a set X (what follows is a transcription of the derivation described in [29]):

i) Start with a set of points X^0, whose points are 0-cells. This corresponds to the basis of the induction.
ii) Build the n-skeleton from the $(n-1)$-skeleton by gluing the n-cells, that we will call e^n_α, using a map $\varphi_\alpha \colon S^{n-1} \to X^{n-1}$. When we use the English word "gluing" actually we mean what in mathematics goes under the theory of

quotient spaces. Therefore to be more specific what we are doing is to define X^n as the quotient space of the disjoint union of X^{n-1} with a collection of n-disks D^n_α (in formula: $X^{n-1} \coprod_\alpha D^n_\alpha$) under the identification of $x \sim \varphi_\alpha(x)$ if $x \in \partial D^n_\alpha$. As a set therefore $X^n = X^{n-1} \coprod_\alpha e^n_\alpha$.

iii) If after n iteration you have obtained the desired space stop and set $X = X^n$.

Notice that what we have called e^n_α are open n-disks, for example in the construction of the torus we used two 1-complexes e^1_1 and e^1_2 that are the two laces a and b of Fig. 1(a), and one 2-complex e^2_1 that is the open disk that forms the surface of the torus.

Now we are ready to define the Euler characteristic.

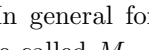
Euler Characteristic. Consider a CW complex X, then the *Euler characteristic* $\chi(X)$ is defined to be the alternating sum

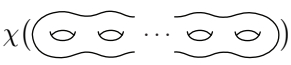
$$\chi(X) = \sum_n (-1)^n c_n, \tag{1}$$

where c_n is the number of n-cells of X.

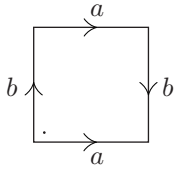
First of all notice that for surfaces of polyhedra the Euler characteristic assumes the familiar form $vertices - edges + faces$; suppose that the surface of such solid has V vertices, E edges and F faces, then it can surely be constructed using V 0-cells, E 1-cells and F 2-cells, moreover since it is a bi-dimensional surface $c_n = 0$ for $n \geq 3$. Then we have $\chi = V - E + F$.

Let us now see how we can apply Eq. (1) to compute some known results. It is immediate that the Euler characteristic of a line is 1, that the one of a circle is 0 and that the one of a disk is 1. To make a sphere we need two 2-cells complexes, therefore the characteristic of a sphere is 2.

From the analysis illustrated in Fig. 1, we see that $\chi($  $) = 0$, while $\chi($  $) = -2$. In general for a closed orientable surface of genus g, the object that earlier we called M_g, we have that it can be built with just one 0-complex, $2g$ 1-complexes and 1 2-complex, therefore we get the following general result

$$\chi\Big(\text{  }\Big) = 2(1-g), \tag{2}$$

where the surface in the previous expression has g handles.



What about some esoteric object like a Klein bottle? Now we show that with the definition given in Eq. (1) it is very easy to compute. All we have to do is to look at the diagram that correspond to the Klein bottle, the one reported at the side of the page; this diagram immediately tells us that the bottle can be built using 1 0-complex, 2 1-complexes and 1 2-complexes ($c_0 = 1$, $c_1 = 2$, $c_2 = 1$ and $c_n = 0$ for $n \geq 3$). Then

$$\chi\Big(\text{  }\Big) = 1 - 2 + 1 = 0. \tag{3}$$

Similar calculations can be done for the Möbious band, the cylinder et cetera.

It is also fairly easy to prove starting from Eq. (1) that for any connected orientable two dimensional surface with h handles and b holes

$$\chi = 2 - 2h - b. \tag{4}$$

The large N model. Now we are ready to consider the large N expansion in a Yang-Mills theory. Let us recall some basic facts and notations that we will use.

The pure Lagrangian of QCD reads

$$\mathscr{L}_{\text{QCD}} = -\frac{1}{4}F_{\mu\nu\,a}F_a^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi, \tag{5}$$

where ψ transforms under gauge transformations with a matrix that belongs to the fundamental representation of $SU(N)$ while $\bar{\psi}$ transforms according to the antifundamental representation:

$$\psi^i \to \psi'^i = U^i_j\psi^j, \qquad \bar{\psi}_i \to \bar{\psi}'_i = U^{*j}_{\ \ i}\bar{\psi}_j, \tag{6}$$

and

$$F^a_{\mu\nu} \equiv \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + iT^a_{bc}A^b_\mu A^c_\nu = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + C^a_{bc}A^b_\mu A^c_\nu, \tag{7}$$

where T^a_{bc} are the generators of $\mathfrak{su}(N)$ in the adjoint representation, and C^a_{bc} are the structure constant of $SU(N)$ ($T^a_{bc} = -iC^a_{bc}$).

The gauge fields, on the other hand, transform according to the adjoint representation of $SU(N)$

$$A_\mu \to UA_\mu U^*, \qquad A_\mu = A^a_\mu t^a, \quad a = 1, \dots, N^2 - 1 \tag{8}$$

Since in order to formulate gauge invariance we need only infinitesimal transformations, we can parametrize $U = e^{i\varepsilon_a t^a}$, and then Eq. (7) and (8) become

$$\begin{aligned} &\delta\psi_i(x) = i\varepsilon_a t^a_{ij}\psi_j(x); \\ &\delta A^a_\mu(x) = i\varepsilon_c T^c_{ab}A^b_\mu(x). \end{aligned} \tag{9}$$

Although Eq. (9) defines the basic transformation properties under the global gauge group, we are interested in the invariance of (6) under transformations induced by the *local gauge group*; this in in fact the reason why we have introduced the fields A^a_μ in the first place through the minimal coupling prescription $\partial_\mu\psi_i \to \partial_\mu\psi_i - it^a_{ij}A^a_\mu\psi_j \equiv D_\mu\psi_i$. Such local transformations are obtained by promoting the transformation parameter ϵ to a smooth function of spacetime; typically we choose $\varepsilon \in \mathscr{D}(\mathbf{R}^4)^{N^2-1}$, or $\mathscr{S}(\mathbf{R}^4)^{N^2-1}$. The Eq. (9) become

$$\begin{aligned} &\delta^\varepsilon\psi_i(x) = i\varepsilon_a(x)t^a_{ij}\psi_j(x); \\ &\delta^\varepsilon A^a_\mu(x) = i\varepsilon_c T^c_{ab}A^b_\mu(x) + \partial_\mu\epsilon^a(x). \end{aligned} \tag{10}$$

As we have already many times remarked the reason why the $1/N$ expansion is significantly harder here is the fact that the fields A^a_μ transforms according to the adjoint representation, this means from a practical point of view that each one of such fields carries two color indices through the index a, or if you prefer A_μ is a $N \times N$ hermitian matrix. So, in a Feynman diagram with gluons you have to

keep track of two indices instead of only one. This means that in diagrams with gluon loops we must learn how to count the number of color loops that are not constrained by the initial and final states, because as we know such loops give rise to factors of N.

In order to start, first of all, as we have done in the ϕ^4 case we need to specify how the large N limit has to be done, that is to say we need to introduce the correct coupling that in this limit can be taken as fixed — the analogue of Eq. 2.1.1–(5). This has been done by 't Hooft in [50] by noticing that the scale of quark masses, that is the physical quantity that we want to keep fixed, does not depend on the coupling g that we can introduce in the Lagrangian Eq. (6) but on the combination $\beta_0 g^2$. To be more specific, it is a very well known result from Renormalization that

$$\Lambda_{\mathrm{QCD}} = \mu e^{1/(2\beta_0 g^2(\mu))}, \tag{11}$$

where β_0 is the first coefficient of the β function (that as you may recall is a scheme-independent quantity) and we know that in the case of $SU(N)$ as the gauge group and with a number of N_f flavors we have

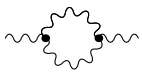
$$\beta_0 = \frac{1}{(4\pi)^2}\frac{11N - 2N_f}{3}. \tag{12}$$

From here you can see that, if we consider the large N limit with g^2N constant, the physical scale Λ_{QCD} is preserved. It seems therefore natural to take

$$g^2 N \equiv \lambda \tag{13}$$

as the 't Hooft coupling. This also means that as N gets bigger $g = O(1/\sqrt{N})$.

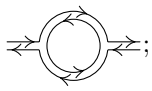
Let us now try to evaluate the order, in powers of N, of some diagrams. Let us start form a low order diagram that contributes to the one-loop gluon vacuum polarization:

 , (14)

the two vertices gives a factor $1/N$, moreover the internal loop has two color indices one of which is fixed by the external legs, while the other is free and gives an additional factor N; this means that the whole diagram is $O(1)$ in N. For this small diagram it was easy to give an argument to derive the order, however in general it is much more convenient to introduce what is known as the *double line notation*, introduced again by 't Hooft in [50]. This notation has been devised in order to keep track of color indices and relies on the already discussed idea that a gluon, with respect to its index structure (and for this purpose only), can be regarded as a quark-antiquark pair, so that it is convenient to represent gluon lines with a pair of fermionic lines with opposite orientations (one line for the quark and one for the antiquark). The cubic, the quartic, and the Yukawa-like vertices can all be expressed in the double line notation just by carefully matching the indices (see [57]) for example the cubic vertex can be expressed as

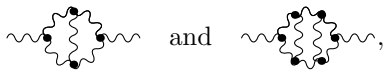
$$(A_\mu)_{ij}(A_\nu)_{jk}(\partial_\mu A_\nu)_{ki}. \tag{15}$$

Following all these rules the diagram in (14) becomes

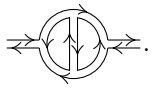
 ; (16)

now the previous discussion becomes clearer and one can see with his own eyes that in fact there is a free color loop in the diagram that gives the factor N.

With this notation it is fairly easy to establish whether a particular diagram survives the large N limit, or if it gives sub-leading corrections to the $1/N$ expansion. For example consider the two diagrams

 , (17)

that contribute respectively to the two and three loop corrections to the gluon vacuum polarization. It is easy to show that both of them are of the same order of (14) since, for example the first one in double line notation it can be redraw as

 . (18)

Then the two more vertices that are introduced are counterbalanced by the additional color loop that appears in (18). Similarly can be argued for the second diagram in (17).

It is not true, however, that all diagrams survive in the large N limit, consider for example the following diagram

 . (19)

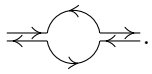
If drawn in double line notation one immediately realizes that it goes like $1/N^2$ because it has six vertices but only one free color loop.

In general one can show that non-planar graphs are suppressed by a factor that is at least $1/N^2$. So experimenting with some more examples one can convince himself that every planar graph made of gluons survives the large N limit and non planar diagrams are suppressed.

What about fermionic loops? Take for example the analogue of (14) with a fermionic loop instead of a gluonic one:

 , (20)

this time we do not have a free internal color loop, therefore the only dependence on N is given by the vertices, this means that the order of the diagram is suppressed by a factor $1/N$ with respect to (16). This argument is made clearer by the double line notation:

 . (21)

This suggests that every time that we replace a gluonic loop with a fermionic one we get a factor $1/N$. Then we conclude that the corrections to the gluon propagator, that we have been discussing so far, have as leading terms the planar graphs with the a minimum number of quark loops.

Actually there is a third prescription that one has to introduce that is relevant when we have fermionic loops coupled with external sources; in fact in that case diagrams with gluons at the edges are suppressed. This means that the leading order in this case is given by diagrams with quarks at the edges.

So far we have discussed, in a somewhat informal way, what are the diagrams that contribute to the leading order of the $1/N$ expansion for the gluonic propagator. We would like now to give a more rigorous discussion for the counting rules for a generic vacuum diagram (with no external lines) like


(22)

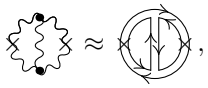
results for a general diagram can be obtained studying what happens when external legs are inserted — for example diagrams with n incoming quarks and n outgoing quarks, which give the n-quarks interaction are suppressed by a factor N^n with respect to the leading vacuum to vacuum diagrams.

In order to facilitate the N counting it is convenient to rescale the Lagrangian (6) as follows

$$\mathscr{L}_{\mathrm{QCD}} = N\left[-\frac{1}{4\lambda}F_{\mu\nu\,a}F_a^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi\right], \tag{23}$$

where λ is the 't Hooft coupling we have just defined.

Because the graphs have no external lines, every index line must close to make an index loop, so a general graph in the double line notation will be made up of a certain number of closed loops. For example


(24)

that has three closed loops.

It is in this context that the previous results on the Euler characteristic becomes handy. In fact let us consider the following construction of a surface X out of a Feynman diagram: Take a general graph (of the kind we are discussing right now); from this graph construct two more graphs, the first one is the usual double line graph (that contains exactly the same information as the graph you have begun with) while the second "structure" graph is obtained just by considering the arcs and the vertices of the original diagram (without distinguishing between fermionic or gluonic propagators nor bothering to write on any arrow). Now regard the structure graph as the 1-skeleton X^1 of the surface that is made up of 0-cells (the vertices) and of 1-cells (the arcs) glued to the vertices.

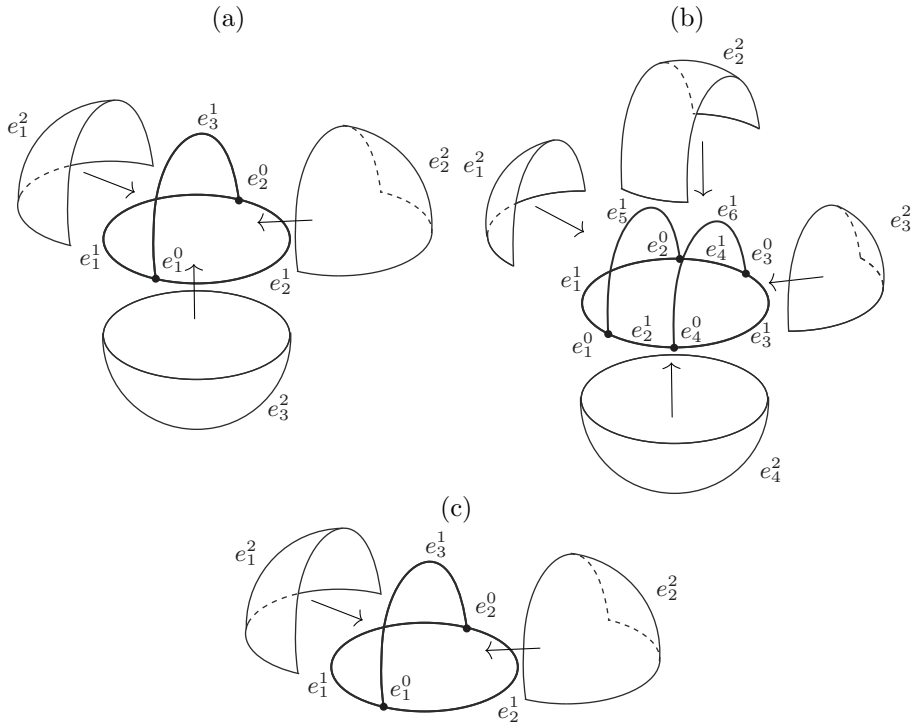


Fig. 2. The structure of a planar vacuum diagram can be considered as a 1-skeleton for a sphere. The gluing instructions φ_α are given by the double arrow notation.

Maybe it is time to take an example to be sure we have control of the procedure we are describing. Suppose we want to build such a surface from the diagram in (24). Then, as we have described we construct the two diagrams

 and  $\equiv X^1$, (25)

as we can see X^1 is made up of the two vertices that we regard as the 0-cells e_1^0 and e_1^0 that form X^0 and of three open arcs, that we can call e_1^1, e_2^1 and e_3^1, glued together in the way that is indicated by the diagram.

So far we have described how to build a one dimensional space, to get a surface we need to describe how to glue 2-cells; and this is where the information provided by the double line graph becomes important. First of all the 2-cells are represented by the interior of the indices loops, for example in our example there are three 2-cells e_1^2, e_2^2, e_3^2. The gluing instructions, that in the previous section we called φ_α are here represented by the arrows. In Fig. 2(a) there is a picture of how the gluing must be done. The same construction for the analogous diagram with two, non crossing, gluons propagator is described in Fig. 2(b).

It is pretty clear that the surfaces that are constructed following this instructions are oriented surfaces, because the 2-cells are glued to the 1-skeleton in a coherent way following the arrows.

In both the examples in Fig. 2 we end up with the same surface: A sphere. Is it possible to construct other spaces starting from a different diagram? The answer is yes, and as we will shortly see the surface that originates a diagram classifies its order in the large N expansion.

A nice example of this can be appreciated by considering this other diagram

(26)

The 1-skeleton obtained from this diagram is the same as in the previous case: X^1. What changes is that now we have only two 2-cells. This changes the gluing instruction that are displayed in Fig. 2(c); and as we can see this time we obtain a hemisphere rather than a sphere, that is to say a sphere with a hole; correspondingly notice that while the previous diagram (the one showed in (25)) is $O(N^2)$ this one is only $O(N)$.

From the previous discussion, and by inspection of the Eq. (23) we immediately see how to generalize the argument: Given a vacuum to vacuum graph with c_0 vertices, c_1 arcs and c_2 index loops it has a factor $N^{c_0-c_1+c_2}$; notice that since we are dealing with surfaces what appears in the exponent is exactly the Euler characteristic that we have defined in Eq. (1), so that the order of a diagram is given by

$$N^{\chi} = N^{2-2h-b}, \tag{27}$$

where we have used Eq. (4). Now since h and b by definition are positive numbers we have proven that the leading order of connected vacuum to vacuum graphs is N^2.

At this point the natural question to ask now is what is the characterization of this diagrams from the point of view of the original theory. If one thinks about this for a minute, referring also to the previous example, it becomes clear that if in the original diagram there is a fermionic loop, in the double line version of the graph this will give rise to an unpaired line that in turn gives an edge of the 1-skeleton to which you don't attach 2-cells, that is to say a hole in the surface.

So the leading diagrams are those with only gluons, otherwise the previous reasoning applies and the Euler characteristic turns out to be lesser than 2.

In meson theory it is relevant to known the leading graphs that contains quarks. These graphs must contain at least one fermionic loop; this means at most are of order N, and since the example in (26) is $O(N)$ we can say that they are of order N. As we have argued before diagrams with external gluon loops are suppressed, so we have our answer: These diagrams are the planar graphs with an external quark loop. The order of any other graph can be derived from Eq. (24).

Let me now summarize the two main results we have proved:

i) The leading connected vacuum to vacuum graphs are the planar graphs made up only with fermions; they are $O(N^2)$.
ii) The leading connected vacuum to vacuum graphs with quarks are the planar graphs with only one quark loop at the external boundary of the graph; they are $O(N)$.

From these two results we can derive some phenomenology on mesons. We will not consider this topic here because is beyond the purpose of this chapter, however excellent references are [57] and [48].

Now that we have shown how to deal with the large N expansion in both vectorial and adjoint representations we will consider the main subject of this chapter: The $\mathbf{C}\mathrm{P}^{N-1}$ model.

I at last deliver to the world a Work which I have long promised,
and of which, I am afraid, too high expectations have been raised.
— JAMES BOSWELL, The Life of Samuel Johnson, LL.D. (1791)

2.2. THE MODEL

In this section we will finally consider the complex projective model, or $\mathbf{C}\mathrm{P}^{N-1}$ model, defined on the whole two dimensional spacetime following the derivations in [56]. Of course many other works on the subject have been published since this article and the others we have mentioned at the beginning of Chapter 1 — notably the work by Di Vecchia, Musto, Nicodemi, Pettorino and Rossi [22] in which the theory has been discussed on the Euclidean lattice in the large N limit, the article by Campostrini and Rossi [11] where the first non leading order in the $1/N$ expansion is worked out or the recent paper by Rossi [45] that considers the large N expansion of the model in the presence of a constant field strength. Many other results, both analytical and numerical, are obtained on the lattice; see for example Campostrini, Rossi [10] and Campostrini, Rossi and Vicari [13, 14].

First of all we will discuss general properties at finite N, the connection of this model (for $N = 2$) with the $O(3)$ model and its descriptions in terms of stereographic coordinates, then in the subsequent section we will completely solve the model at large N showing some interesting properties that this theory shares with QCD.

2.2.1. The finite N model

In order to introduce the model for a generic N it is convenient to start with the so called $O(3)$ sigma model (see [48]).

The theory can be described in terms of the three components of a vector field $u(x) = (u^1(x), u^2(x), u^3(x))$, subjected to the constraint that u must lie on a two-dimensional sphere of radius r of the internal space. Such constraint can be expressed as usual by

$$|u|^2 \equiv \sum_i (u^i)^2 = r^2, \tag{1}$$

this means that whatever the dynamics of the model is, it will describe a motion on the surface of the sphere. Such dynamics can be specified by the Lagrangian

$$\mathscr{L} = \frac{1}{2}\partial_\mu u \cdot \partial^\mu u; \tag{2}$$

notice that the coupling of the theory can be considered to be the inverse of the radius r, so that classically one can choose $r = 1$, however from a quantum mechanical point of view r is a parameter that undergoes renormalization.

The Lagrangian in Eq. (2) can be regarded as the only Lagrangian compatible with $O(3)$ and Lorentz invariance with the lowest number of derivatives, and that can be constructed without the aid of any other tensor.

As we may have noticed till now we haven't said anything about the dimension of spacetime, so that in general the action of such a model is

$$S = \frac{1}{2}\int d^D x\, \partial_\mu u \cdot \partial^\mu u, \tag{3}$$

then some comments are due on the dimensionality of spacetime. As usual $[u] = (D-2)/2$, then from Eq. (1) we have that $[r^2] = 2[u] = D-2$. As we have remarked the coupling is the inverse of r, and this means that its dimensionality is $2-D$. The theory is then renormalizable if $D \leq 2$.

Besides this, the model exhibits non-trivial soliton-like solutions only in two spacetime dimensions; this is enough to convince ourselves to study the model in the case $D = 2$. However if we make this assumption we may also want to add to the Lagrangian the so called θ term that arises only in two dimensions (for what concerns this theory) and is compatible with the $O(3)$ invariance. Such term assumes the form

$$\mathscr{L}_\theta = \frac{\theta}{8\pi} u^a (\partial_\mu u^b)(\partial_\nu u^c)\varepsilon^{\mu\nu}\varepsilon_{abc}, \tag{4}$$

where μ and ν range from 1 to 2 while a, b and c are internal indices that assume values $1, 2, 3$. The two ε terms are Levi-Civita tensors.

The curious fact about this term is that classically it does not modify the theory since it does not affect the equations of motion. In order to verify this assertion we only need to verify that the first variation of $S_\theta[u] \equiv \int d^2x\, \mathscr{L}_\theta$ with respect to the variation of $u \to u + \delta u$ is zero. In fact we have that

$$\delta S_\theta[u] = \frac{\theta}{8\pi}\int d^2x\, \varepsilon^{\mu\nu}\varepsilon_{abc}[2\partial_\mu(u^a \delta u^b \partial_\nu u^c) + 3(\delta u^a \partial_\mu u^b \partial_\nu u^c)], \tag{5}$$

the first term between square brackets is a total derivative, so it is vanishing. What can we say about the second term? Remember that the field u is constrained on a sphere, then both the variations δu and the derivatives $\partial_\mu u$ are perpendicular to u, this means that for example $\varepsilon_{abc}\delta u^a \partial_\mu u^b$ is a vector $v_{\mu\, c}$ that is parallel to u_c, and consequently perpendicular to $\partial_\nu u^c$, then the scalar product $v_{\mu\, c}\partial_\nu u^c$ vanishes. This concludes the proof that $\delta S_\theta[u] = 0$.

Although this term is non relevant classically, it is still relevant quantum mechanically, so it is important to acknowledge its compatibility with the symmetry of the theory. However we will not deal with it here.

Complex coordinates. We started with the $O(3)$ sigma model because we can show his equivalence with the first complex projective space $\mathbf{C}\mathrm{P}^1$ just by performing a very clever change of coordinates.

The key observation here is the fact that we are using three fields coordinates plus a constraint to parametrize a target space that is two dimensional: A sphere.

There is, on the other hand, a standard mathematical procedure to represent the points of a sphere through complex numbers. This procedure is called stereographic projection.

The basic idea is that we can project every point of a sphere on a plane from, for example, the north pole. Consider the setting displayed in Fig. 1: A

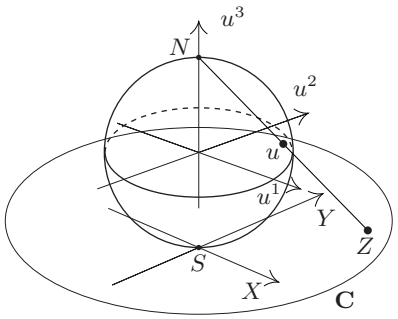


Fig. 1. Stereographic projection of a point u from the north pole N of a sphere of radius r to a point Z of the complex plane $\mathbf{C}$. The two coordinate systems have been chosen so that u^1 has the same orientation as X and u^2 the same as Y.

sphere of radios r lays on the complex plane touching it with the south pole; then consider a point u on the sphere with coordinates (u^1, u^2, u^3), such point is mapped into the point $Z \equiv X + iY$, that is the projection of u from the north pole. The real and imaginary parts of Z can be easily evaluated from geometric considerations starting from Fig. 1 and turn out to be

$$X = \frac{2u^1}{1 - u^3/r}, \qquad Y = \frac{2u^2}{1 - u^3/r}, \tag{6}$$

this means that

$$Z = 2\frac{u^1 + iu^2}{1 - u^3/r}. \tag{7}$$

Transformations (6) are invertible everywhere except at the points $u^3 = r$ (the north pole). Then if we assume $u^3 \neq r$ we can invert them, using the constraint $(u^1)^2 + (u^2)^2 + (u^3)^2 = r^2$ to obtain the inverse maps that send a point in the complex plane on the sphere. Some algebra shows that

$$u^1 = \frac{X}{1 + \dfrac{X^2 + Y^2}{4r^2}}, \quad u^2 = \frac{Y}{1 + \dfrac{X^2 + Y^2}{4r^2}}, \quad u^3 = \frac{r(X^2 + Y^2) - 4r^3}{4r^2 + X^2 + Y^2}, \tag{8}$$

or if you want in terms o the complex variable Z Eq. (8) becomes

$$u^1 = \frac{\Re Z}{1 + Z^*Z/4r^2}, \quad u^2 = \frac{\Im Z}{1 + Z^*Z/4r^2}, \quad u^3 = \frac{rZ^*Z - 4r^3}{4r^2 + Z^*Z}. \tag{9}$$

With this transformations at hand we can rewrite the Action (3) in terms of the unconstrained field $Z(x)$. With some simple but lengthy calculations we can show that

$$S = \frac{1}{2}\int d^2x \, \frac{\partial_\mu Z^*(x)\partial^\mu Z(x)}{\left(1 + \dfrac{Z^*(x)Z(x)}{4r^2}\right)^2}. \tag{10}$$

In this representation the $O(3)$ sigma model is known as $\mathbf{C}\mathrm{P}^1$ model. But before talking about the higher dimensional (of the target space) generalization of this model, that is the one we are interested in, it is worth spending a few words on how the symmetries under rotation of the original model are realized here. First of all notice that (10) is invariant under

$$Z \to e^{i\alpha} Z, \tag{11}$$

that is a rotation of the point Z of an angle α. Looking at the stereographic projection in Fig. 1 we can immediately understand this kind of symmetry and the fact that it is inherited from the symmetry rotations around the axis u^3 of the original model. The remaining part of the original symmetry group, that is generated by rotations around u^1 and u^2 cannot be guessed immediately, however studying what happens to the projected point after such rotations one can derive the following transformations that are a symmetry of (10):

$$\delta X = \epsilon + \epsilon^* X^2, \qquad \delta X^* = \epsilon^* + \epsilon X^{*2}, \tag{12}$$

where ϵ is a complex parameter with small modulus.

As this brief introduction to $\mathbf{C}\mathrm{P}^1$ may suggests, $\mathbf{C}\mathrm{P}^{N-1}$ models can be defined in a similar "geometric" way using the theory of Kähler manifolds. However we won't follow this idea and instead we will use the more direct procedure of using the definition of a complex projective space to implement the symmetries of such space directly into a Lagrangian formulation.

First of all then we have to be a little more specific on what a complex projective space is.

Complex Projective Spaces. First of all let us start with *real* projective spaces. Consider vectors in $\mathbf{R}^{n+1}$ and introduce the following equivalence relation: We say that two vectors v and w are equivalent ($v \sim w$) if they differ for a non-null real constant λ ($v = \lambda w$). The equivalence classes in which $\mathbf{R}^{n+1} \setminus \{0\}$ breaks into are the elements of the real projective space $\mathbf{R}\mathrm{P}^n$, then as a space $\mathbf{R}\mathrm{P}^n = (\mathbf{R}^{n+1} \setminus \{0\})/\sim$.

There is actually a very nice way to visualize the real projective space. By definition a line through the origin in $\mathbf{R}\mathrm{P}^{n+1}$ is a point in $\mathbf{R}\mathrm{P}^n$; any of such lines intersects the unit sphere $S^n \subset \mathbf{R}^{n+1}$ in two diametrically opposite points, then $\mathbf{R}\mathrm{P}^n$ can also be thought as a sphere S^n in which diametrically opposite points have been glued together.

One can prove that $\mathbf{R}\mathrm{P}^n$ is not just a set of points but it has the structure of a smooth manifold (see [34]).

From the definition we have given of the real projective space it is clear that it admits a generalization to the complex field $\mathbf{C}$. Instead of considering real $(n+1)$ dimensional vectors, let us consider the space of $n+1$ dimensional complex vectors $\mathbf{C}^{n+1} = \{\, (z_1, \ldots, z_{n+1}) \mid z_1, \ldots, z_{n+1} \in \mathbf{C} \,\}$, and define $\mathbf{C}\mathrm{P}^n$ as the space of lines through the origin of $\mathbf{C}^{n+1}$. To be more specific, as before introduce an equivalence relation $\bowtie$ in $\mathbf{C}^{n+1}$ so that $\nu \bowtie \omega$, with ν and ω in $\mathbf{C}^{n+1}$, if $\nu = \lambda\omega$ with $\lambda \neq 0$ and $\lambda \in \mathbf{C}$. Then define $\mathbf{C}\mathrm{P}^n = (\mathbf{C}^{n+1} \setminus \{0\})/\bowtie$. Again

just as before one can think of $\mathbf{C}\mathrm{P}^n$ as a S^{2n+1} sphere with the identification of two points that differ for a global phase, for example ν and $e^{i\alpha}\nu$. Moreover $\mathbf{C}\mathrm{P}^n$ is a $2n$ dimensional manifold (the topology is the quotient topology inherited from the natural projection $\pi\colon \mathbf{C}^{n+1} \setminus \{0\} \to \mathbf{C}\mathrm{P}^n$), that can be equipped with a smooth structure ([34]).

The Lagrangian. The $\mathbf{C}\mathrm{P}^{N-1}$ Lagrangian is built with a multiplet of complex fields $n(x) = (n^1(x), \dots, n^2(x))$. Each component of n is a scalar (complex) field and, as a multiplet transforms according to the fundamental representation of $SU(N)$ (just like the quarks).

The interesting property of such a multiplet is the fact that the target space of $n(x)$ is taken to be $\mathbf{C}\mathrm{P}^{N-1}$, this actually means that the "length" of n must be fixed

$$|n|^2 = \sum_i n_i^* n_i = r, \tag{13}$$

this implements the first constraint of $\mathbf{C}\mathrm{P}^{N-1}$, that is to say the fact that $n \in S^{2n-1}$. However we need to introduce also a way to identify phase connected field configurations: That is to say two field configurations n and n' related by the relation

$$n'(x) = e^{i\alpha(x)} n(x). \tag{14}$$

This constraint can be implemented in a Lagrangian formulation by means of an auxiliary gauge field A_μ that guaranties a $U(1)$ *local* invariance.

We are now ready to write down the Lagrangian of our model. The $U(1)$ local invariance through the field $A_\mu(x)$ is implemented in the Lagrangian by means of the covariant derivative $D_\mu \equiv \partial_\mu - iA_\mu$, while the constraint on the modulus of n can be made explicit with the introduction of a Lagrange multiplier $\lambda(x)$. Let us see this in details; consider the Lagrangian

$$\mathscr{L} = (D_\mu n_i)^* D^\mu n_i - \lambda(n_i^* n_i - r), \tag{15}$$

then (13) is recovered as the equation of motion $\partial\mathscr{L}/\partial\lambda = 0$ for the non-propagating field $\lambda(x)$. While the invariance under (14) of the Lagrangian can be made explicit using the equation of motion for A_μ:

$$A_\mu = \frac{i}{2r}[(\partial_\mu n_i^*) n_i - n_i^* \partial_\mu n_i], \tag{16}$$

and inserting it back in the Lagrangian one obtains a nontrivial self interaction. Notice that if one does that substitution all the interaction terms are proportional to $1/r$ since A_μ is proportional to $1/r$ and the term that is quadratic in A is multiplied by $n_i^* n_i = r$, then as before the "size" of $\mathbf{C}\mathrm{P}^{N-1}$, r, can be regarded as the parameter that describes the interactions of the theory and it is usually expressed in the literature in terms of the coupling constant g^2 through the convenient combination

$$r = \frac{4\pi}{g^2}. \tag{17}$$

It is worth observing that the description we gave here of $\mathbf{C}\mathrm{P}^1$ is equivalent to the one we have derived in Eq. (10); in fact in our new formalism the model is described by a two components complex vector $n = (n^1, n^2)$, but if we identify $Z(x)$ with the ratio $n^1(x)/n^2(x)$ the model assumes the form in Eq. (15).

We are now ready to solve this model in the large N limit.

2.2.2. Solution at large N

In this theory, as for QCD, we will see that the 't Hooft coupling is determined by the combination of g^2 and N that appears in the physical scale of the problem, and we will see that such combination is g^2N.

The Action of the theory, as it can be read off from (15) is

$$S[n, n^*, A, \lambda] = \int dt\, dx\, [(D_\mu n_i)^* D^\mu n_i - \lambda(n_i^* n_i - r)]. \tag{1}$$

As we have done in the case of the $1/N$ expansion in the ϕ^4 model (remember Eq. 2.1.1 –(13)) we will be able to write an effective Action for the fields A and λ. To do this we introduce, as usual, the partition function of the theory in the Minkowsky space:

$$\begin{aligned} Z(r) &= \int [dn_i\, dn_i^*][dA_\mu]\, \delta_F(|n|^2 - r) \exp\left[i \int d^2x\, (D_\mu n_i)^* D^\mu n_i \right] \\ &= \int [dn_i\, dn_i^*][dA_\mu][d\lambda]\, e^{iS[n, n^*, A, \lambda]}, \end{aligned} \tag{2}$$

where δ_F is the functional delta function.

Since the Action (1) is quadratic in the n fields the expression of the partition function can be reduced by performing an integration over the fields n, n^*. To do this first of all it is necessary to rewrite the action as a quadratic form in order to be able to use standard formulae for gaussian integration.

Integration by parts shows us that

$$S[n, n^*, A, \lambda] = n_i^*(x)\, \mathscr{A}_{ix,jy} n_j(y) + r \int d^2x\, \lambda, \tag{3}$$

where $\mathscr{A}_{ix,jy} = [(i\partial + A)^2 - \lambda]\delta^{(2)}(x - y)\delta_{ij}$ and we have used the summation convention both on discrete and continuous indices. At this point remember that for complex variables, gaussian integrals give the following result

$$\int [dn_i\, dn_i^*] e^{i n_i^*(x)\, \mathscr{A}_{ix,jy} n_j(y)} = \frac{1}{\det(i\,\mathscr{A}/2\pi)} \tag{4}$$

Because $\mathscr{A} = A \otimes \mathbb{1}_N$, then $\det(\mathscr{A}) = (\det A)^N$ where $A_{x,y} = [(i\partial + A)^2 - \lambda]\delta^{(2)}(x - y)$, moreover the determinant can be conveniently expressed as an exponential via $\det A = \exp[\mathrm{Tr} \log A]$. All these relations yield

$$\frac{1}{\det(iA/2\pi)} = \mathcal{N} \exp\Big[- N \int d^2x\, \log[(i\partial_x + A(x))^2 - \lambda(x)] \Big]. \tag{5}$$

Therefore the generating functional is

$$Z(r) = \int [dA_\mu][d\lambda]\, \exp\Big[- N \operatorname{Tr} \log[(i\partial + A)^2 - \lambda] + ir \int d^2x\, \lambda(x)\Big]; \qquad (6)$$

as we have already remarked the interpretation provided in Eq. 2.1.1–(13) also applies here if we set

$$S_{\text{eff}}[\lambda, A] = iN \operatorname{Tr} \log[(i\partial + A)^2 - \lambda] + \int d^2x\, r\lambda(x), \qquad (7)$$

as the effective action for the fields λ and A.

Vacuum and Renormalization. Let us study the theory at large N. To do this we can use the saddle point approximation. Translational invariance fixes the vacuum expectation value of λ to be a constant, while Lorentz invariance gives $\langle A_\mu \rangle = 0$.

The saddle point equation $\delta S_{\text{eff}}/\delta\lambda = 0$ is

$$iN \operatorname{Tr} \left(\frac{1}{\partial^2 + \lambda} \right) + r = 0; \qquad (8)$$

and in order to evaluate the trace one must sum all the eigenvalues of the operator that in momentum space are just $1/(-p^2 + \lambda)$. Then the so called gap equation becomes

$$iN \int \frac{d^2p}{(2\pi)^2}\, \frac{1}{-p^2 + \lambda} + r = 0. \qquad (9)$$

Notice that the integral that appears in Eq. (9) is logarithmically divergent. Then we have to regularize the expression introducing an ultraviolet cutoff Λ in the integral (dimensional regularization is discussed in [12]). Moreover it is convenient to perform a Wick rotation $p^0 \to ip^0$, so that the integrand will have imaginary poles:

$$r = N \int \frac{d^2p}{(2\pi)^2}\, \frac{1}{p^2 + \lambda}, \qquad (10)$$

where now p^2 is the square modulus of the euclidean four dimensional vector (p^1, p^2, p^3, p^4). Switching to polar coordinates, the integral on the disk of radius Λ becomes

$$r = \frac{N}{4\pi} \int_0^\Lambda dp\, \frac{2p}{p^2 + \lambda} = \frac{N}{4\pi} \log \frac{\Lambda^2 + \lambda}{\lambda} = \frac{N}{4\pi} \log \frac{\Lambda^2}{\lambda}, \qquad (11)$$

since Λ is a ultraviolet cutoff ($\Lambda \gg \sqrt{\lambda}$).

To renormalize the radius of $\mathbf{C}\mathrm{P}^{N-1}$ we must therefore introduce in our model a counterterm proportional to λ, to be more specific

$$\mathscr{L} \to \mathscr{L} + \lambda \log \frac{\Lambda^2}{\mu^2}, \qquad (12)$$

where μ is as usual an arbitrary energy scale that is needed to subtract the divergency. With this subtraction the renormalized radius becomes

$$r_R = \frac{N}{4\pi} \log \frac{\mu^2}{\lambda}, \tag{13}$$

and then the renormalized coupling $\alpha_R \equiv g_R^2$ is just

$$\alpha_R = \frac{(4\pi)^2}{N \log(\mu^2/\lambda)}. \tag{14}$$

At this point the α beta function can be computed directly from Eq. (14) yielding

$$\beta_\alpha \equiv \mu \frac{d\alpha_R}{d\mu} = -\frac{2N}{(4\pi)^2} \alpha_R^2. \tag{15}$$

It is a standard result that one can easily derive from the definition of the beta function that the quantity $\mu[\exp - \int_0^{\alpha_R} d\alpha/\beta_\alpha]$ (see [3]) is a Renormalization Group invariant, for example in QCD this constant is what is called $\Lambda_{\rm QCD}$ and what sets an energy scale for the theory. Here if we call such a constant $\Lambda_{\rm CP}$ we immediately have from Eq. (15) and from the definition just given that

$$\Lambda_{\rm CP} = \mu e^{-(4\pi)^2/2N\alpha_R}, \tag{16}$$

however algebraic manipulations of Eq. (14) show that

$$\lambda = \mu e^{-(4\pi)^2/N\alpha_R}, \tag{17}$$

then we have the interesting result that $\sqrt{\lambda} = \Lambda_{\rm CP}$. This clearly suggests that in the large N limit the λ field becomes a mass term for the n fields. In fact inspection of Eq. 2.2.1–(15) shows that the generated mass for n is $m \equiv \sqrt{\lambda}$; this mass also coincides with the scale of the theory:

$$m = \Lambda_{\rm CP}. \tag{18}$$

Some comments are now due. As we have anticipated the large N limit that we have performed must be taken keeping $g^2 N$ fixed as one can see from Eq. (16). Moreover as we have promised Eq. (14) shows that the theory shares with QCD the property of asymptotic freedom.

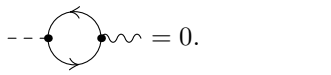
Spectrum and Confinement. In order to derive the spectrum of the theory we must consider fluctuations around the vacuum solution that we have just derived. In the formalism of the functional integral this means that we have to consider correction to the saddle point approximation, expanding the effective action (7) around the saddle point. To do this we must have $\lambda = 0$ at the vacuum, so we must shift $\lambda \to \lambda - \langle\lambda\rangle$.

If we perform such expansion we realize that linear terms vanish by definition, and by counting the powers of N as we did in the previous examples one can prove that the cubic and higher order terms in λ and A_μ are suppressed at least by a factor $1/\sqrt{N}$.

Then we must only worry about the quadratic terms that are obtained by $\delta^2 S_{\text{eff}}/\delta\lambda^2$, $\delta^2 S_{\text{eff}}/\delta\lambda\delta A_\mu$ and $\delta^2 S_{\text{eff}}/\delta A_\mu \delta A_\nu$.

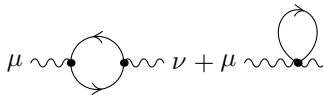
The quadratic term in λ is non-vanishing, but as remarked in [56] it does not give significant contribution to the confinement mechanism and so we will ignore the correction to λ and we will just retain the saddle point approximation $\lambda = m^2$.

The mixing term vanishes, diagrammatically this means that



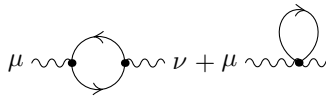
$$= 0. \tag{19}$$

Then the only term that we need to compute is the quadratic term in A_μ. Diagrammatically this term corresponds to



$$+ \qquad ; \tag{20}$$

of course one can evaluate this sum directly, however notice that gauge invariance requires this sum to be $(-\eta_{\mu\nu}p^2 + p_\mu p_\nu)$ up to some function $f(p^2)$. Moreover we know that those diagrams are proportional to N (since they have an unconstrained n loop), and that the multiplicative function $f(p^2)$ is non-singular at $p = 0$ since the particles n_i have acquired mass (then their propagator is regularized in the IR). This means that we can always make an expansion in powers of p^2 of the function $f(p^2)$ around $p^2 = 0$: $f(p^2) = f(0) + O(p^2)$. Putting all this together one has that



$$+ \qquad = N(-\eta_{\mu\nu}p^2 + p_\mu p_\nu))(f(0) + O(p^2)). \tag{21}$$

Since we are interested in the particle structure of the theory we only care about low energy oscillations around the vacuum, this means that we can forget the $O(p^2)$ corrections.

Notice now that $-\eta_{\mu\nu}p^2 + p_\mu p_\nu$ is noting else but the standard gauge field kinetic term $(-1/4)(F_{\mu\nu})^2$ written in momentum space. This actually means that the A_μ field, that has been introduced as a non-dynamical auxiliary field, now dynamically acquires a kinetic term.

Let us now see that this result in two dimensions implies confinement of the n fields. Instead of using the full Lagrangian of the theory, in what follows we just need to consider an effective theory which has the features of the model we are studying: A mass for the n fields and a kinetic term for A_μ.

The effective Lagrangian then will be of the form

$$\mathscr{L}_{\text{eff}} = (D_\mu n)^* D^\mu n - m^2 n_i^* n_i - \frac{N f(0)}{4}(F_{\mu\nu})^2, \tag{22}$$

moreover it is convenient to rescale the photon field A_μ in order to have a standard normalization of the kinetic term, to do this the correct rescaling is

$\sqrt{Nf(0)}A_\mu \to A_\mu$; doing so the effective Lagrangian assumes the standard form

$$\mathscr{L}_{\text{eff}} = -\frac{1}{4}F_{\mu\nu}^2 + (\nabla_\mu n)^* \nabla^\mu n - m^2 n_i^* n_i, \tag{23}$$

where now the covariant derivative is $\nabla_\mu = \partial_\mu - i(Nf(0))^{-1/2}A_\mu$. Notice that it is now clear that the n fields acquire a charge that is $1/\sqrt{Nf(0)}$; since we are in the large N limit actually we have that the fields n_i are weakly charged.

However, in one space dimension the Coulomb potential is linear, this means that no matter how small the coefficient is it will confine charges. In fact the potential between a quark (n_i) at the space point x and an antiquark (n_i^*) at the space point y is

$$V(x,y) = Nf(0)|x-y|. \tag{24}$$

This property is the one that we would like to prove in QCD. Notice that in this model the proof is quite straightforward once we work in the large N limit.

CHAPTER THREE

COMPLEX PROJECTIVE MODEL ON A STRIP

I had absolutely no idea that what I'd foreseen as a sort of "salad course"
would eventually turn out to be the main dish.

— DONALD E. KNUTH, *Combinatorial Algorithms* (2011)

Round his brow he had a peculiar yellow band, with brownish speckles,
which seemed to be bound tightly round his head.
As we entered he made neither sound nor motion.
"The band! the speckled band!" whispered Holmes.

— SHERLOCK HOLMES, in *The Adventure of the Speckled Band* (1892)

That was the origin of Buckland, a thickly inhabited strip between
the river and the Old Forest, a sort of colony from the Shire.

— J.R.R. TOLKIEN, *The Fellowship of the Ring* (1954)

In the previous chapter we have discussed the complex projective model defined on a two-dimensional spacetime. As we remarked this model has been studied because of its affinity with QCD and, at the same time, because it could be solved exactly in the large N limit; unlike Yang-Mills theories.

The present chapter, is devoted the discussion of the same model on a spatially finite domain of length L; the fields are defined on a worldstrip $[0\,..\,L]\times\mathbf{R}_t$. This model will be discussed in the large N limit as well.

The study of this system is motivated, as we have already discussed in Section 1.2.4, by the research in confinement, and by the results that has been achieved in the dual Meissner effect involving nonabelian strings and monopoles: The internal moduli dynamics of a vortex that ends into two monopoles is exactly the complex projective space on a strip with the length of the vortex string.

The model has been taken under examination, for similar reasons, notably by Shifman, Monin, and Yung [40] who studied the same system in a compactified strip (a cylinder), by Milekhin [38] with Dirichelt boundary conditions, by Dunne and Ünsal [24, 25] with twisted boundary conditions and eventually by Bolognesi, Konishi and Ohashi [9] that reconsidered the analysis carried on by Milekhin in more details. (Recently the study of boundary effects from a more general prospective has been proposed in [4].)

An interesting analysis of this model on a finite spacetime volume is given in [2]; in this work it is discussed the θ-dependence in relation to how the large N limit and the large volume limit are performed.

In what follows first of all we will present the problem, and we will make some preliminary discussions on boundary conditions, then we will find the effective action in which the original degrees of freedom are reduced just in the same way we did in Section 2.2.2.

In the following section we attack the problem using the large N limit and we derive the analogue of gap equation Eq. 2.2.2–(8). We then show how those gap equations can be solved in the case of periodic boundary condition (i.e. when the translational invariance ansaz of the vacuum still holds).

The main part of the chapter is then to consider the case of more general boundary conditions such as Dirichler or Neumann conditions, and to solve the gap equations under those conditions. Numerical analysis of the problem is presented along with some analytical discussions of the property of the system.

The problem of the energy is finally addressed. The main result here is the fact that we are able to prove the finiteness of the energy density and to study the large L behavior of the total energy.

Since we will have to handle many formulae which are valid only in the interval $[0\,..\,L]$ we will sometimes use a very useful *bracket notation*

$$[\text{statement}] = \begin{cases} 1, & \text{if the statement is true;} \\ 0, & \text{if the statement is false.} \end{cases} \tag{1}$$

When using this notation the Kronecker delta can be rewritten as $\delta_{ij} = [i = j]$. Because conventions like (1) were introduced by K.E. Iverson in 1962, (1) is often called *Iverson's convention*.

Theory is the first term in the Taylor series of practice.
— THOMAS M. COVER (1992)

3.1. THE MODEL ON A FINITE INTERVAL

CONSIDER the system defined by the Lagrangian in Eq. 2.2.2–(15), but this time suppose that we force the domain of the fields to be the rectangular, infinite domain $S = [0\,..\,L] \times \mathbf{R}_t$. As always we need to specify the behavior of the fields at the boundaries of S. Two sides of the rectangular region are at infinity; on those regions we will assume that the fields are vanishing. On the other two sides we have to impose some other kind of conditions.

In the following we will be concerned mainly with Neumann and Dirichlet conditions, as well as periodic ones. Let us be more specific since we have to remember that the n field must also satisfy a constraint on its norm.

The Lagrangian of the theory is specified by the integral of a Lagrangian density over the domain S. Then it is convenient to use Iverson's notation to write the Lagrangian density:

$$\mathscr{L} = \Big((D_\mu n_i)^* D^\mu n_i - \lambda(n_i^* n_i - r) \Big) [x \in S], \tag{1}$$

however since we have to impose boundary conditions on fields $n_1, \ldots, n_N$ that are already constrained to have $n_i^*(x) n_i(x) = r$ for every $x \in S$, it is much easier to separate the set of the n fields into a field $\sigma \equiv n_1(x)$ and the remaining set of fields $n_2, \ldots n_N$; in this way Dirichlet and Neumann conditions can be expressed as

$$\sigma(0,t) = \sigma(L,t) = \sqrt{r}, \qquad n_i(0,t) = n_i(L,t) = 0 \quad \text{for} \quad i = 2, \ldots, N; \tag{2}$$

$$\partial_x n_i(0,t) = \partial_x n_i(L,t) = 0 \quad \text{for} \quad i = 1, \ldots, N. \tag{3}$$

In what follows we will refer to the first set of conditions as *DD* conditions while the second set will be referred as *NN* conditions.

For periodic conditions we simply identify the $(0,t)$ points of S with the corresponding (L,t) points.

After these considerations is then possible to rewrite the Lagrangian in Eq. (1) separating the first component σ from the others, so that manipulations will be easier. The Lagrangian that from now on will be considered is then:

$$\mathscr{L} = \Big(\sum_{j=2}^{N} [(D_\mu n_j)^* D^\mu n_j - \lambda(n_j^* n_j - r)] + (D_\mu \sigma)^* D^\mu \sigma - \lambda |\sigma|^2 \Big) [x \in S]. \tag{4}$$

The partition function of the theory in Minkowsky space this time can be written as

$$Z(r) = \int [d\varphi] \exp\Big(i \int d^2x\, \mathscr{L} \Big), \tag{5}$$

where we have introduced the shorthand notation $[d\varphi]$ for $[dn_i^* dn_i][d\sigma][dA_\mu][d\lambda]$. Notice also that because of the $U(1)$ invariance 2.2.2–(14) we can make the

field $\sigma(x)$ real, simply performing such transformation with the angle $\alpha(x) = -\arctan \Im\sigma/\Re\sigma$.

As we did in Section 2.2, we integrate out the n fields — that appear quadratically into the action — but not the σ field since it does not have vanishing boundary conditions. The resulting partition function will then describe an effective theory for the fields σ, λ and A_μ.

Basically the procedure is the same as the one we have already used: We just need to be a little bit careful about boundary conditions. Remember that we want to discuss periodic, *DD* and *NN* conditions; however we can, for the moment, avoid breaking up the discussion for each one of these cases since, as we will show, the process of integrating out the fields is the same for all this three choices of boundary conditions.

Let us rewrite the n-dependent part of the Lagrangian as a quadratic form. In doing so one needs to do integration by parts both for the term $\partial_\mu n_i^* \partial^\mu n_i$ and for the term $\partial_\mu n_i^* n_i A_\mu$; this time however one also has to consider boundary contributions coming from the integration by parts. Fortunately they all vanish for all boundary conditions that we have considered. In fact the boundary term of $\partial_\mu n_i^* \partial^\mu n_i$ is $n_i^* \partial^\mu n_i|_0^L$ that is vanishing for both *DD* or *NN* conditions (of course also for periodic condition). The same reasoning goes for the other boundary term. This means that the structure of the quadratic form is the same as in Eq. 2.2.2–(3) except that this time the matrix $\mathscr{A}$ is a $(N-1)\times(N-1)$ matrix with respect to the internal index i. Also you have to add to the expression of $\mathscr{A}$ a factor $[x \in S]$ that we won't write down explicitly.

With this warnings clear in mind we can now finally perform the gaussian integration over the n^*n fields. The result can be expressed in terms of the effective action:

$$S_{\text{eff}} = i(N-1)\,\mathrm{Tr}\log[(i\partial + A)^2 - \lambda] + \int d^2x\,\{D_\mu\sigma D^\mu\sigma - \lambda(\sigma^2 - r)\}[x \in S]. \quad (6)$$

Again notice that the fact that the system is defined on a strip affect the way the trace must be taken: We have to trace an operator that acts on functions defined on $[0\,..\,L]$ with some specified boundary conditions. As we will see this will complicate the analysis for the gap equations.

Since we are ultimately interested in the large N limit from now on we will not make distinction between N and $N-1$, and whenever $N-1$ appears we will always replace it with N.

3.1.1. Gap Equations

We are now ready to study the gap equations of this theory whose solutions are the vacuum fields, just like in the model defined on the whole plane.

This equations have already been studied in some articles notably by Shifman [40] and Milekhin [38], but in those articles they considered specific translationally invariant boundary conditions (Shifman) or they assumed it (Milekhin). We will, on the other hand, follow the derivation of Bolognesi, Konishi and Ohashi ([9]) in which they carefully study the complete gap equations.

The whole point of the matter is that with respect to the previous case now we cannot assume translational invariance since the system is defined on a strip, and by definition translation of the spacial coordinate is not a symmetry of the system: Given a point we can always tell how far it is from the borders of the strip, but this distance changes if we translate the point.

On the other hand Lorentz invariance is still a good invariance of the system and therefore we can still assume $\langle A_\mu \rangle = 0$. This will simplify our, already tricky, discussion a lot.

So summing up, we cannot assume that λ and σ are constant (unless we are only interested in the case with periodic boundary conditions, in which case translational invariance is recovered), but we can still assume that the vacuum expectation value of A_μ is vanishing.

Sturm Liouville problem. As we will see, in order to write down the gap equations and for many other further manipulations it will be useful to study the Sturm Liouville problem

$$-f_n''(x) + \lambda(x) f_n(x) = \omega_n^2 f_n(x), \tag{1}$$

with various boundary conditions together with the first order variation of the eigenfunctions f_n and the eigenvalues ω_n^2 under small variations of the potential $\lambda(x)$.

Let us, before doing anything else, recall some basic features of a Sturm Liouville problem (see [41] and [16]). The general form in which it is usually stated is to find the solution of the differential equation

$$\frac{1}{\rho(x)} \left[\frac{d}{dx} \left(p(x) \frac{d}{dx} \right) - q(x) \right] u + \lambda u = 0, \tag{2}$$

where $u = u(x)$, $p(x)$, $q(x)$ and $\rho(x)$ are regular enough functions on a finite interval $[\alpha \,..\, \beta]$. Furthermore suppose that $u(x)$ must satisfy some boundary conditions (for what follows we will assume *DD* condition, however the results are the same with periodic or *NN* conditions)

$$u(\alpha) = u(\beta) = 0. \tag{3}$$

For example in the case of interest described in Eq. (1) we have $\rho(x) \equiv 1 \equiv p(x)$, $q(x) = \lambda(x)$.

Suppose also that $\rho(x)$ is a positive function, then we can define a scalar product

$$(f, g)_\rho = \int_\alpha^\beta f^*(x) g(x) \rho(x)\, dx, \tag{4}$$

and with respect to this scalar product it is easy to check that two eigenfunctions u_1 and u_2 relative to different eigenvalues are orthogonal.

A more interesting mathematical result is the fact that the eigenvalues of this problem are all real and positive, better yet one can prove that the eigenvalues are strictly increasing

$$\lambda_1 < \lambda_2 < \lambda_3 < \cdots, \tag{5}$$

with $\lambda_n \to \infty$ as $n \to \infty$. Also the eigenfunctions $\{u_n\}$ are an orthogonal complete set.

This results are readily applicable to the Eq. (1), plus in that case the eigenfunctions can be taken to be real and orthonormal:

$$(f_n, f_m) \equiv \int_0^L f_n(x) f_m(x)\, dx = \delta_{nm}. \tag{6}$$

The completeness relation is usually stated as

$$\sum_n f_n(x) f_n(y) = \delta(x-y), \quad \text{for any } x \text{ and } y \text{ in } [0\,..\,L], \tag{7}$$

this assures us that any regular enough function can be written in $[0\,..\,L]$ as a Fourier series of such eigenfunctions. Notice however that the completeness relation in Eq. (7) does not automatically satisfy the boundary conditions; for example take $y = 0$ in Eq. (7), then this reduces to $\sum_n f_n(x) f_n(0) = \delta(x)$. If we assume *DD* boundary conditions we immediately see that the completeness relation, stated in this way is not consistent with these conditions.

As we will now show (as it is known also from electrostatic) it is possible to modify such completeness relation in such a way that the sum remains the same in the interval $[0\,..\,L]$ but is smartly modified outside this region in such a way so that it becomes consistent with boundary conditions. Such method is called the method of the *mirror images*, and of course it is slightly different for *DD* , *NN* and periodic conditions.

Let us focus for the moment on the problem with *DD* boundary conditions, and consider instead of Eq. (7) the following relation

$$\sum_n f_n(x) f_n(y) = \sum_k \big(\delta(x-y+2kL) - \delta(x+y+2kL)\big), \tag{8}$$

(for *NN* conditions just replace the minus sign between the two delta functions with a plus sign, while for periodic conditions the right hand side of the equation would be $\sum_k \delta(x-y+kL)$) then, as we promised, if both x and y are in $[0\,..\,L]$ (or in $[0\,..\,L)$ for periodic conditions) Eq. (8) reduces to Eq. (7) because only the first δ in the term $k = 0$ is non-vanishing in this case. Let us see this in details. Consider first of all the term $k = 0$, this is made up of two terms $\delta(x-y)$ and $\delta(x+y)$, since both x and y are between 0 and L it cannot happen that $x = -y$, then the only surviving term is $\delta(x-y)$. Now for every other term different from $k = 0$, as one can readily convince himself by trying to compute a couple of them, one always find δs that impose the condition $x = y + c$ where $y + c \notin [0\,..\,L]$.

So why it is convenient to consider such a modified completeness relation? The reason become clear when, as before, one tries to evaluate $\sum_n f_n(x) f_n(0)$; we have

$$\sum_n f_n(x) f_n(0) = \sum_k \big(\delta(x+2kL) - \delta(x+2kL)\big) = 0, \tag{9}$$

the same thing holds for *NN* conditions. So we will take Eq. (8) as our completeness relation.

The other relevant question that we have to ask ourselves about the f_ns and the ω_ns is how do they change when the potential λ changes a little. This problem is the same that we have to address in standard time-independent perturbation theory in Quantum Mechanics.

The problem can be stated as follows. Consider Eq. (1) and perform a small variation of the potential λ: $\lambda(x) \to \lambda(x) + \delta\lambda(x)$. As a consequence of this change, the eigenvalues and the eigenfunctions will change correspondingly: Call δf_n the change in the eigenfunctions, and $\delta\omega_n^2$ the change in the eigenvalues. In general we will have that

$$\begin{aligned} \delta\omega_n^2 &= \delta_1\omega_n^2 + \delta_2\omega_n^2 + \cdots ; \\ \delta f_n(x) &= \delta_1 f_n(x) + \delta_2 f_n(x) + \cdots , \end{aligned} \tag{10}$$

where $\delta_i\omega_n^2$ and $\delta_i f_n$ are both $O(\epsilon^i)$, where $\epsilon = O(\delta\lambda)$. Our goal is to find the first order variations of ω_n^2 and $f_n(x)$, namely $\delta_1\omega_n^2$ and $\delta_1 f_n(x)$.

Let us start by finding $\delta_1\omega_n^2$. By definition

$$-f_n''(x)-\delta_1 f_n''(x)+(\lambda(x)+\delta\lambda(x))(f_n(x)+\delta_1 f_n(x)) = (\omega_n^2+\delta_1\omega_n^2)(f_n(x)+\delta_1 f_n(x)), \tag{11}$$

then using Eq. (1) and rearranging the terms we can write

$$\delta_1 f_n'' + \lambda\delta_1 f_n + f_n\delta\lambda = \omega_n^2\delta_1 f_n + f_n\delta_1\omega_n^2. \tag{12}$$

The next step is to project this equation onto f_n using the scalar product (6); notice that by requiring that $f_n + \delta f_n$ must be properly normalized we obtain, up to the first order in ϵ, that $1 = (f_n + \delta f_n, f_n + \delta f_n) = 1 + 2(f_n, \delta_1 f_n)$; then $(f_n, \delta_1 f_n) = 0$, now it is a straightforward calculation to show that

$$\delta_1\omega_n^2 = (f_n^2, \delta\lambda) = \int_0^L f_n^2(x)\delta\lambda(x)\,dx. \tag{13}$$

For the $\delta f_n(x)$s, start back again from (12) but this time project this equation onto $f_m(x)$ with $m \neq n$. After some integrations by parts one gets

$$(f_m, \delta\lambda f_n) = \delta_{nm}\delta_1\omega_n^2 + (\omega_n^2 - \omega_m^2)(f_m, \delta_1 f_n), \tag{14}$$

and since we have assumed that $m \neq n$, and because we know that the eigenvalues are non degenerate (that is to say it cannot happen that $\omega_m = \omega_n$ unless $m = n$) dividing both sides by $(\omega_n^2 - \omega_m^2)$ we have

$$(f_m, \delta_1 f_n) = \frac{(f_m, \delta\lambda f_n)}{\omega_n^2 - \omega_m^2}. \tag{15}$$

Finally from the fact that $\{f_n\}$ is a complete set in $[0\,..\,L]$, we eventually obtain

$$\delta_1 f_n = \sum_{m\geq 1}(f_m, \delta_1 f_n) f_m = \sum_{\substack{m\geq 1 \\ m\neq n}} f_m \frac{(f_m, \delta\lambda f_n)}{\omega_n^2 - \omega_m^2}. \tag{16}$$

With this work done we are now ready to derive the gap equations.

The Equations. The gap equations, or saddle point equations, this time (as we will see shortly) form a coupled differential system of two equations:

$$\begin{cases} \dfrac{\delta S_{\text{eff}}}{\delta \sigma}[\lambda, \sigma] = 0; \\ \dfrac{\delta S_{\text{eff}}}{\delta \lambda}[\lambda, \sigma] = 0. \end{cases} \tag{17}$$

The first equation can be immediately read off from the effective action 3.1–(6), since the complicated trace factor does not contains σ, and it gives simply $\sigma''(x) - \lambda(x)\sigma(x) = 0$ with $x \in [0\,..\,L]$. Notice that σ is a function of space only since the vacuum, even thought it is no more invariant under space translations, still retains the invariance under time translations.

On the other hand it is not as easy to express the first equation as a differential equation for λ and σ, in fact we will not be able to write down a nice equation in which the λ dependence is explicit, but it will appear only through the eigenvalues and eigenfunctions of the differential operator $-d^2/dx^2+\lambda(x)$ defined on the interval $[0\,..\,L]$ that of course depend on λ. Things are also complicated by the fact that the trace must be taken on the spectrum of functions with specific boundary conditions.

So the first thing that we have to do, in order to obtain reasonable equations is to rewrite the trace-term in Eq. 3.1–(6); to be more precise we need to express this term as a function of the eigenvalues of $-d^2/dx^2 + \lambda(x)$. Let us see how this can be done.

It will be convenient to evaluate the derivative with respect to λ of the trace term that appears in the effective action:

$$\frac{\partial}{\partial \lambda} \operatorname{Tr} \log(-\partial^2 - \lambda) = \operatorname{Tr} \left(\frac{1}{\partial^2 + \lambda} \right). \tag{18}$$

Then the trace can be obtained integrating the result with respect to λ.

Now the trick is to Fourier transform with respect to both time and space: A generic function $F(x,t)$ that satisfies boundary conditions 3.1–(2) and 3.1–(3) can be decomposed as

$$F(x,t) = \int_{-\infty}^{\infty} dp^0 \, e^{ip^0 t} \sum_n a_n(p^0) f_n(x), \tag{19}$$

with f_n being the complete set of eigenfunctions for a given class of boundary conditions that are the solutions of the Sturm-Liouville problem that we have just discussed. Still all this discussions equally applies to *DD* , *NN* , as well as to periodic boundary conditions.

Fourier transforming Eq. (18) one immediately obtains

$$\frac{\partial}{\partial \lambda} \operatorname{Tr} \log(-\partial^2 - \lambda) = \frac{T}{2\pi} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} dp^0 \, \frac{1}{-(p^0)^2 + \omega_n^2}, \tag{20}$$

where $T/2\pi$ is the usual phase space factor.

The integral over p^0 can be computed explicitly so that the derivative of the trace gives $i(T/2)\sum_{n=1}^{\infty}\omega_n^{-1}$. To obtain the final result we need to integrate over λ. The λ dependence of ω_n can be understood from Eq. (1): $\omega_n = \sqrt{\lambda - f_n''/f_n}$, so we eventually discover that

$$\mathrm{Tr}\log(-\partial^2 - \lambda) = iT\sum_{n\geq 1}\omega_n. \tag{21}$$

Using the fact that the fields are actually constant in the vacuum, and Eq. (21) that we have just derived, the effective action 3.1–(6) becomes

$$S_{\text{eff}} = -T\Big(N\sum_{n=1}^{\infty}\omega_n + \int_0^L dx\,[(\sigma'(x))^2 + \lambda(\sigma^2 - r)]\Big). \tag{22}$$

With the action written in this form we are now in position to compute the second equation in (17); in fact since we have that $\delta\omega_n = \delta\omega_n^2/2\omega_n$ with $\delta\omega_n^2$ given to the first order by Eq. (13) under variation of λ

$$\delta S_{\text{eff}} = -TN\sum_{n\geq 1}\frac{(f_n^2, \delta\lambda)}{2\omega_n} - T\int_0^L(\sigma^2 - r)\delta\lambda\,dx \tag{23}$$

from which one immediately derives the gap equation for λ

$$N\sum_{n\geq 1}\frac{f_n^2(x)}{2\omega_n} + \sigma^2(x) - r = 0. \tag{24}$$

Before starting to analyze these gap equations, and try to solve them with various boundary conditions, we will show that in fact it exists a very nice representation of the term $\sum_n f_n^2/2\omega_n$ that appears in Eq. (24) in terms of the propagator of the n fields. Such representation will also give us extra insight on what really the second gap equation we have derived means.

The n-fields propagator. An elegant way to evaluate the propagator for the n fields is to use the functional integral method. Let us see this standard derivation in some details. We will make this derivation in Euclidean space since manipulations with the functional integrals are more sound within this framework, and it won't make any difference for the usage that we will do of this result.

We won't bother to change our symbols so, from now on, in this paragraph, t will be the Euclidean time (what in literature is sometimes called t_E or τ). First of all we need to introduce a generating functional for the correlations functions of the n-fields:

$$Z[J, J^*] = \int [dn_i^* dn_i]\exp\big\{-I_n + [n_i^*(x,t)J_i(x,t) + n_i(x,t)J_i^*(x,t)]\big\}, \tag{25}$$

where I_n is the Euclidean action for the n fields namely $I_n = \sum_{j=2}^{N}(\partial_\mu n_j^*\partial_\mu n_j + \lambda n_j^* n_j)[0 < x < L]$, and the sum over μ gives the standard scalar product in $\mathbf{R}^4$. Moreover J and J^* are the coupled currents to n^* and n respectively.

The sums here are on both discrete and continuous indices and the sum over i has to be performed starting from 2. Notice also that Z is a functional of the auxiliary field λ (and then the propagator will depend on λ. This seems reasonable since the auxiliary field is like a position-dependent term).

Then the propagator by definition is

$$\langle n_i^*(x,t) n_j(y,s)\rangle = \frac{1}{Z[0,0]} \left.\frac{\delta^2 Z}{\delta J_i(x,t)\delta J_j^*(y,s)}\right|_{J=J^*=0}. \tag{26}$$

In order to be able to use this expression (in a constructive way) we must carry on the explicit form of Z; such a computation can be done since the integral over n is gaussian.

As we have previously done in the computations for the effective action we need to rewrite the action as a quadratic form. To achieve this the derivative term must be integrated by parts:

$$\begin{aligned}\partial_\mu n_i^*(x,t)\partial^\mu n_i(x,t)[0<x<L] &= -n_i^*(x,t)\partial_\mu(\partial^\mu n_i(x,t)[0<x<L]) \\ &= -n_i^*(x,t)\partial^2 n_i(x,t)\cdot[0<x<L],\end{aligned} \tag{27}$$

since $[0<x<L]' = [x=L] = \delta(x-L)$, the action can be rewritten as

$$\begin{aligned}I_n &= n_i^*(x,t)[0<x<L](-\partial^2+\lambda)\delta(x-y)\delta(t-s)\delta_{ij} n_j(y,s) \\ &\equiv n_i^*(x,t)\,\mathscr{M}_{ixt,jys} n_j(y,s).\end{aligned} \tag{28}$$

Now we can apply the well known formula for the gaussian complex integral to obtain:

$$Z[J,J^*] = \mathcal{N} e^{-J_i^*(x,t)\left(\mathscr{M}\right)^{-1}_{ixt,jys} J_j(y,s)}, \tag{29}$$

where $\mathcal{N}$ is a term that does not depend on J and J^* (but depends on λ), and using equation (26) we have that

$$\langle n_i^*(x,t) n_j(y,s)\rangle = \left(\mathscr{M}\right)^{-1}_{ixt,jys}. \tag{30}$$

Therefore if we call $\mathscr{D}_{ij}(x,t;y,s)$ the Euclidean propagator, we have just proved that $\mathscr{D}_{ij}(x,t;y,s) = \left(\mathscr{M}\right)^{-1}_{ixt,jys}$, and then it can be computed by solving the differential equation

$$(-\partial^2+\lambda(x))\,\mathscr{D}_{ij}(x,t;y,s) = \delta_{ij}\delta(t-s)\delta(x-y), \tag{31}$$

with corresponding boundary conditions at $x,y=0$ and $x,y=L$ (DD or NN depending on the choice we make on the n fields). However, as we have done for the f_ns, such boundary conditions can be automatically satisfied if instead of solving (31) one consider the following differential equation

$$(-\partial^2-\lambda(x))\,\mathscr{D}_{ij}(x,t;y,s) = \delta_{ij}\delta(t-s)\sum_k \left(\delta(x-y+2kL)\mp\delta(x+y+2kL)\right), \tag{32}$$

where, as we discussed for Eq. (8) the $\mp$ term refers to the boundary conditions that one chooses (DD or NN).

The claim is that the solution of such equation is given by

$$\mathscr{D}_{ij}(x,t;y,s) \equiv \delta_{ij} \sum_{n\geq 1} \frac{e^{-\omega_n|t-s|}}{2\omega_n} f_n(x) f_n(y). \tag{33}$$

This result gives us an handy representation for the sum-term in the second gap equation: Define $\mathscr{D}_\epsilon(x) \equiv \mathscr{D}_{ii}(x,t+\epsilon;x,t)$, with ϵ a positive number then we clearly have that

$$\lim_{\epsilon\to 0} \mathscr{D}_\epsilon(x) = \lim_{\epsilon\to 0} \sum_{n\geq 1} \frac{e^{-\epsilon\omega_n}}{2\omega_n} f_n^2(x) = \sum_{n\geq 1} \frac{f_n^2(x)}{2\omega_n}. \tag{34}$$

This derivation also suggests that $\mathscr{D}_\epsilon(x)$ at finite ϵ is a good regularization for the propagator. Moreover Eq. (24), using the representation that we have just established, can be cast in the form

$$\sum_{i=1}^{N} \langle n_i^*(x,t) n_i(x,t)\rangle - r = 0 \tag{35}$$

since we know that $\sum_{i=1}^N \langle n_i^*(x,t)n_i(x,t)\rangle = \sum_{i=2}^N \langle n_i^*(x,t)n_i(x,t)\rangle + \sigma^2(x)$, and we have just shown that the sum in the second member of the previous expression is just the sum that appears in Eq. (24).

Let us stop for a moment to contemplate how nice this result is. Observe that Eq. (35) is the quantum version of the constraint $|n|^2 = r$ that defines the model. This is the quantization condition on the size of $\mathbf{C}\mathrm{P}^{N-1}$ and it is here recovered as a gap equation.

Mind the gap!

— Audible warning of the Underground

3.1.2. Solution of Gap Equations

In the previous section we derived the two gap equations that define the vacuum structure of the $\mathbf{C}\mathrm{P}^{N-1}$ model on a strip. This two equations written together are:

$$\begin{cases} \sigma''(x) - \lambda(x)\sigma(x) = 0; \\ N \displaystyle\sum_{n\geq 1} \frac{f_n^2(x)}{2\omega_n} + \sigma^2(x) - r = 0, \end{cases} \tag{1}$$

we have also worked out an alternative version of the second gap equation, given by Eq. 3.1.1–(35) that will be useful in what follows.

The discussion has been carried on in such a way that we have considered the various possible boundary conditions of our interest together. In discussing the solutions to such equations as we have already remarked, we have to distinguish between the translational invariant conditions, that is to say the periodic conditions, and the *DD* and *NN* conditions.

First of all we will show that Eq. (1) are compatible with periodic boundary conditions, under the assumption of translational invariance. We will then see

that they are incompatible with *DD* and *NN* boundary conditions under the same assumption.

In the next section we will carry on the analysis for *DD* and *NN* conditions without the assumption of translational invariance.

For all the subsequent analysis it turns out to be convenient to split the second equation in (1) into two: We separate the average of this equation from the fluctuations around the mean. Let us be more precise. Consider the second equation in (1), call the first member of such equation $A(x)$, the equation then is $A(x) = 0$. The trivial observation is that you can always write such equation as $A(x) - (1/L)\int_0^L A(x)\,dx + (1/L)\int_0^L A(x)\,dx = 0$; now by definition we have that $A(x) - (1/L)\int_0^L A(x)\,dx$ has zero average, this means that if we average another time this equation we get $(1/L)\int_0^L A(x)\,dx = 0$, and inserting this back into the original equation we obtain $A(x) - (1/L)\int_0^L A(x)\,dx = 0$. So

$$A(x) = 0 \Longleftrightarrow \begin{cases} \dfrac{1}{L}\displaystyle\int_0^L A(x)\,dx = 0 \\ A(x) - \dfrac{1}{L}\displaystyle\int_0^L A(x)\,dx = 0 \end{cases}, \tag{2}$$

if $A(x)$ is an integrable function. Applying this reasoning to our equation we have — once we call $\sigma^2(x) - (1/L)\int_0^L \sigma^2(x)\,dx \equiv \tilde{\sigma}^2(x)$ — that

$$N\sum_{n\geq 1}\frac{f_n^2(x)}{2\omega_n} + \sigma^2(x) - r = 0 \Longleftrightarrow \begin{cases} \dfrac{N}{L}\displaystyle\sum_{n\geq 1}\frac{1}{2\omega_n} + \frac{1}{L}\int_0^L \sigma^2(x)\,dx - r = 0; \\ N\displaystyle\sum_{n\geq 1}\frac{1}{2\omega_n}\left(f_n^2(x) - \frac{1}{L}\right) + \tilde{\sigma}^2(x) = 0. \end{cases} \tag{3}$$

We will refer to the first equation as the constant part of the lambda gap equation, while the second can be referred to as the varying part.

3.1.2.1. Translational Invariant Solutions. Suppose now that the vacuum is invariant under spacial translations. Then the same argument we used in the two dimensional $\mathbf{C}\mathrm{P}^{N-1}$ in the previous chapter applies here, and tells us that both λ and σ are constants. As we did in 2.2 let us call $\lambda = m^2$.

First of all let us see if Eq. 3.1.2–(1) are consistent with this assumption once we assume *Dirichlet boundary conditions*. In this case we know exactly the solutions to the Sturm-Liouville problem 3.1.1–(1) since λ is a constant:

$$f_n(x) = \sqrt{\frac{2}{L}}\sin\frac{n\pi x}{L}, \qquad \omega_n^2 = \left(\frac{n\pi}{L}\right)^2 + m^2 \quad n = 1, 2, \ldots . \tag{1}$$

Instead of considering directly Eq. 3.1.2–(1) it is more convenient to find an explicit solution of Eq. 3.1.1–(32); then we need to find the solution of

$$(-\partial^2 + m^2)\,\mathscr{D}(x, t; y, s) = \delta(t - s)\delta(x - y), \tag{2}$$

where we have dropped the indices i and j defining $\mathscr{D}_{ij} = \mathscr{D}\delta_{ij}$. The solutions to such equation are the modified Bessell function K_0; in fact we can regard Eq. (2) as the defining equation for those functions. We prefer, however, to define K_0 through a series and then prove that such series satisfies Eq. (2). We won't discuss here the convergence properties of such series; this being said let us define

$$\begin{aligned} K_0(z) = -\left(\log\frac{z}{2}+\gamma\right) I_0(z) + \frac{1/4\,z^2}{(1!)^2} + \left(1+\frac{1}{2}\right)\frac{(1/4\,z^2)^2}{(2!)^2} \\ + \left(1+\frac{1}{2}+\frac{1}{3}\right)\frac{(1/4\,z^2)^3}{(3!)^2} + \dots\,, \end{aligned} \tag{3}$$

where in turn I_0 is defined by the expansion

$$I_0(z) = 1 + \frac{1/4\,z^2}{(1!)^2} + \frac{(1/4\,z^2)^2}{(2!)^2} + \frac{(1/4\,z^2)^3}{(3!)^2} + \dots\,. \tag{4}$$

(For such expressions see [1 Eq. 9.6.12 and 9.6.13].) Then putting all together we obtain

$$K_0(z) = \sum_{k\geq 0}(-\log z/2 - \gamma + H_k)\frac{(1/4\,z^2)^k}{(k!)^2}, \tag{5}$$

where H_k are the *Harmonic numbers* defined by (for a detailed discussion on such numbers see Don Knuth *Fundamental Algorithms* [30])

$$H_k = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k} = \sum_{n=1}^{k}\frac{1}{n}. \tag{6}$$

With this expressions at hand one can verify that the solution to Eq. (2) is

$$\mathscr{D}(x,t;y,s) = \frac{1}{2\pi}K_0(m\sqrt{(x-y)^2+(t-s)^2}) \tag{7}$$

The superposition principle provide us with the solution of the full equation 3.1.1–(32): We eventually have that

$$\begin{aligned} \mathscr{D}(x,t;y,s) = \frac{1}{2\pi}\sum_k \Big[& K_0(m\sqrt{(x-y+2kL)^2+(t-s)^2}) \\ & - K_0(m\sqrt{(x+y+2kL)^2+(t-s)^2})\Big]. \end{aligned} \tag{8}$$

At this point the idea is to use the propagator representation explained in Eq. 3.1.1–(34) to rewrite gap equations 3.1.2–(3). In particular we have that

$$\frac{N}{L}\sum_{n\geq 1}\frac{1}{2\omega_n} = \frac{N}{L}\lim_{\epsilon\to 0}\int_0^L \mathscr{D}_\epsilon(x)\,dx; \tag{9}$$

if we combine this with the hypothesis of constancy of σ^2, the constant part of the lambda gap equation then becomes

$$\frac{N}{L}\lim_{\epsilon\to 0}\int_0^L \mathscr{D}_\epsilon(x)\,dx + \sigma^2 - r = 0, \tag{10}$$

where $\mathscr{D}_\epsilon(x)$ is given in an explicit form in Eq. (8). Now in order to take the limit as ϵ goes to zero we need to study the behavior of $\mathscr{D}_\epsilon$ around zero. We have that

$$\mathscr{D}_\epsilon(x) = \frac{1}{2\pi}\sum_k \Big[K_0(m\sqrt{4(kL)^2+\epsilon^2}) - K_0(m\sqrt{4(x+kL)^2+\epsilon^2})\Big]. \qquad (11)$$

Notice that, for a generic k, in the second term, as ϵ becomes smaller and smaller $K_0(m\sqrt{4(x+kL)^2+\epsilon^2})$ is $K_0(2m|x+kL|)+O(\epsilon^2)$. For the first term things are not as simple as in the second term because in the sum is included a divergent term; the term $k=0$ is $K_0(m\epsilon)$ and looking at the expansion in Eq. (5) we see that $K_0(m\epsilon) = -\log m\epsilon/2 - \gamma + O(\epsilon^2)$. This term diverges as a logarithm as ϵ approaches 0. For all other terms, however, the limit is finite.

Then let us rewrite the regularized version of Eq. (10) that is obtained just by dropping the limit $\epsilon \to 0$. Because of the considerations we have just made it is convenient to separate the zeroth term in the first sum, so that $\sum_k K_0(m\sqrt{4(kL)^2+\epsilon^2}) = K_0(m\epsilon) + \sum_{k\neq 0} K_0(m\sqrt{4(kL)^2+\epsilon^2})$; notice also that in the sum on $k\neq 0$ the positive and negative terms contributes in the same way, then this sum is just $2\sum_{k\geq 1} K_0(m\sqrt{4(kL)^2+\epsilon^2}) = 2\sum_{k\geq 1} K_0(2mkL) + O(\epsilon^2)$. Putting all together we end up with the following equation

$$\frac{N}{2\pi}\Big(-\log\frac{m\epsilon}{2} - \gamma + 2\sum_{k\geq 1} K_0(2mkL)\Big) - \frac{N}{2\pi L}\sum_k \int_0^L K_0(2m|x+kL|)\,dx + \sigma^2 - r = O(\epsilon^2); \qquad (12)$$

now we clearly see that the gap equation is divergent; however we can regularize the expression just considering the equation at finite ϵ.

We can compute explicitly the sum of the integrals once we notice that, after a change of variables $x \to x - kL$ the k dependence goes into the extrema of integration, so that the sum becomes $\sum_k \int_{kL}^{(k+1)L} K_0(2m|x|)\,dx = \int_{-\infty}^{\infty} K_0(2m|x|)\,dx = \pi/2m$. Eq. (12) can then be simplified to

$$\frac{N}{2\pi}\Big(-\log\frac{m\epsilon}{2} - \gamma + 2\sum_{k\geq 1} K_0(2mkL)\Big) - \frac{N}{4mL} + \sigma^2 - r = O(\epsilon^2). \qquad (13)$$

Here the scale of the theory $\Lambda_{\rm CP}$ can be introduced by requiring consistency of Eq. (13) with Eq. 2.2.2–(18). This can be done by taking the limit $L\to\infty$ of Eq. (13) and substituting in the resultant expression $m = \Lambda_{\rm CP}$. Let us do this; from the large L limit since $\lim_{z\to\infty} K_0(z) = 0$, and $\sigma^2 = 0$ in this limit, we have

$$\frac{N}{2\pi}\Big(-\log\frac{\epsilon\Lambda_{\rm CP}}{2} - \gamma\Big) - r = 0, \qquad (14)$$

where we have already done the substitution $m = \Lambda_{\mathrm{CP}}$. This equation tells us that here Λ_{CP} is defined by

$$\Lambda_{\mathrm{CP}} = 2\epsilon^{-1} \exp\left(-\frac{8\pi^2}{Ng^2} - \gamma\right) \tag{15}.$$

So far everything seems to be consistent, since also the first of Eq. 3.1.2–(1) seems to be satisfiable just by taking $\lambda = 0$. (Indeed in Milekin's article this was indicated as one of the two possible phases that the system can take, the other being the one in which $\sigma = 0$.)

However things are not so easy since one must remember that Eq. (10) is just the constant part of the λ gap equation, as one can see from Eq. 3.1.2–(3); actually as we will now show problems arise from the second "varying" part.

Since σ is a constant $\tilde{\sigma} = 0$ by definition; so rewritten with the aid of the propagator representation (Eq. 3.1.1–(34)) the other equation we must consider is exactly

$$\lim_{\epsilon\to 0}\left(\mathscr{D}_\epsilon(x) - \frac{1}{L}\int_0^L \mathscr{D}_\epsilon(x)\,dx\right) = 0 \tag{16}$$

that, using the relations we have just computed, in turn gives

$$\frac{N}{4mL} - \frac{N}{2\pi}\sum_k K_0(2m|x - kL|) = 0. \tag{17}$$

Clearly this equation cannot be satisfied for all x, to actually see this explicitly consider what happens for $x = 0$ (or $x = L$). At those points the sum contains just one divergent term, but since $K_0(z)$ is a positive function this divergence cannot be absorbed by the infinite terms of the series, therefore the equation is inconsistent.

As expected *DD* boundary conditions are not compatible with the invariance assumption. What about *NN* conditions?

For Neumann boundary conditions the solutions to Eq. 3.1.1–(1) are

$$f_n(x) = \sqrt{\frac{2}{L}}\cos\frac{n\pi x}{L}, \qquad \omega_n^2 = \left(\frac{n\pi}{L}\right)^2 + m^2 \quad n = 1, 2, \ldots . \tag{18}$$

For what concerns the propagator representation the only thing that changes is the sign in Eq. (11), and this has no other effect on Eq. (17) than that of changing the overall sign.

$$-\frac{N}{4mL} + \frac{N}{2\pi}\sum_k K_0(2m|x - kL|) = 0. \tag{19}$$

Then is clear that also *NN* conditions are not compatible with translational invariance.

Now, let us show that, instead, for periodic boundary conditions such equations are consistent.

In this case, in fact, a complete set of solutions of the Sturm Liouville problem 3.1–(1) is given by the sum of the set in Eq. (1) and that in Eq. (18). In

this case the varying part of the gap equation is given by the sum of Eq. (17) and Eq. (19) and then it is automatically satisfied. The other part of the equation will give the solution of the problem (such solutions has been studied in details in [40] and they will be partially discussed by us in Section 3.2.1).

3.1.2.2. General Solutions. We will now try to solve the full gap equations 3.1.2–(1) without the restricting hypothesis of invariance under space translations. Even though gap equations cannot be solved analytically we will understand their behavior at the boundaries and we will find solutions using numerical methods.

Before getting into this kind of analysis it is necessary to review the WKB approximation since it will be extensively used especially to approximate the form of eigenvalues and eigenfunctions of the high energy modes.

WKB approximation. In Quantum Mechanics when a particle has sufficiently high momentum, it will be described by a wave function that varies very rapidly, much more than the potential in which it lives. Since Schödinger equation can be solved exactly for a constant potential, the idea here is that one can solve it approximately for a slowly varying one.

The method that is commonly used has been introduced (in three independent papers in 1926) by Gregor Wentzel, Hendrik Kramer and Leon Brillouin, from where the name WKB.

Consider the Schrödinger equation in one dimension

$$u''(x) + k^2(x)u(x) = 0, \tag{1}$$

in Quantum Mechanics $k(x)$ is related to the potential $U(x)$ through the relation $k(x) = \sqrt{(2m/\hbar)(E - U(x))}$ if the particle has mass m. If one assumes U to be enough regular (actually it is sufficient to be continuous) then we can divide the real axis $\mathbf{R}$ into two regions V and Λ with $V = \{\, x \in \mathbf{R} \mid U(x) > E \,\}$ and define Λ as $\Lambda = \{\, x \in \mathbf{R} \mid U(x) < E \,\}$; V and Λ are known as the classically allowed and classically forbidden region respectively. The points in $\mathbf{R}$ that don't belong to neither V or Λ are called classical turning points. Of course when $x \in V$ we have that $k(x)$ is pure imaginary, on the other hand when $x \in \Lambda$, $k(x)$ is real.

Now if $k(x)$ were constant the solution of Eq. (1), as it is well known, is a linear combination of $\exp(ikx)$ and $\exp(-ikx)$. Then we might guess that when $k(x)$ is slowly varying the solutions are of the form

$$A(x) \exp\Big(i \int k(x)\, dx\Big), \quad \text{or} \quad A(x) \exp\Big(- i \int k(x)\, dx\Big). \tag{2}$$

In order to see if this guess is consistent, and in order to understand better what does *slowly varying* mean let us substitute this solutions into Eq. (1), and see under which conditions this equation is satisfied.

If we do so we obtain a differential equation for $A(x)$:

$$A''(x) \pm 2ikA' \pm ik'A = 0, \tag{3}$$

in order to be able to solve this we have to make the further assumption that $A''(x)$ can be neglected, actually what we are saying here is that we assume that $|A''| \ll 1$ and then we will verify that in fact this is the case if $k(x)$ is slowly varying in a sense that we will specify in a moment.

If we drop the term A'' in Eq. (3), then this equation becomes $2kA' + k'A = 0$. This can be integrated, and the solution is $A(x) = C/\sqrt{k(x)}$, where C is the integration constant. What is A''? And under which hypothesis can it be ignored? By direct calculation we see that

$$A''(x) = \frac{C}{2}\left(-\frac{k''}{k^{3/2}} + \frac{3k'}{2k^{5/2}}\right); \tag{4}$$

this means that we can ignore the term A'' in Eq. (3), or better saying $|A''| \ll |k'A|$, if

$$\left|\frac{k''}{k'}\right| \ll k, \quad \text{and} \quad \left|\frac{k'}{k}\right| \ll k. \tag{5}$$

Under this conditions we have proved that solutions to Eq. (1), in the classically allowed region is a linear combination of

$$\frac{1}{\sqrt{k(x)}}\exp\Big(i\int k(x)\,dx\Big) \quad \text{and} \quad \frac{1}{\sqrt{k(x)}}\exp\Big(-i\int k(x)\,dx\Big). \tag{6}$$

On the other hand if we are in the classically forbidden region Λ the solution is a linear combination of

$$\frac{1}{\sqrt{\kappa(x)}}\exp\Big(\int \kappa(x)\,dx\Big) \quad \text{and} \quad \frac{1}{\sqrt{\kappa(x)}}\exp\Big(-\int \kappa(x)\,dx\Big), \tag{7}$$

and still this time this is a good approximation if κ satisfies the exact same conditions written in (5).

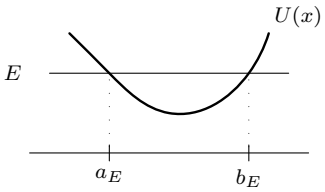


Fig. 1. Potential.

As it is known the tricky part of this approximation are the matching conditions at the turning points. Remember what happens in a typical quantum mechanical system where we have $V = (a_E\,..\,b_E)$ and $\Lambda = (-\infty\,..\,a_E) \cup (b_E\,..\,\infty)$ as the one pictured in Fig. 1. WKB approximation is valid only away from the turning points a_E and b_E; the problem is that the solution at plus and minus infinity determines the correct combination of solutions in (6) to be used in the subset of $(a_E\,..\,b_E)$ where WKB can be used; and because they are not contiguous regions one must study what happens near the turning points. There Eq. (1) can be solved directly making a linear approximation of the potential, and under the condition that there is an overlapping region between the range of validity of such solution and WKB approximation, namely if

$$\frac{2\sqrt{2mU'(b_E)}}{3\hbar}\left(\frac{2U'(b_E)}{|U''(b_E)|}\right)^{3/2} \gg 1, \tag{8}$$

we can find matching conditions (for further details on this well known analysis see [33] or [55] on which we based our notations).

For example in the example illustrated in Fig. 1, the matching conditions performed at b_E imply that the correct combination of solutions (6) is

$$u(x) = C_b k^{-1/2} \cos\Big(\int_x^{b_E} k(y)\,dy - \frac{\pi}{4}\Big), \tag{9}$$

while matching conditions around a_E give

$$u(x) = C_a k^{-1/2} \cos\Big(\int_{a_E}^{x} k(y)\,dy - \frac{\pi}{4}\Big), \tag{10}$$

where C_a and C_b are two constants. In order for these two solutions to be compatible it necessary that

$$\int_{a_E}^{b_E} k(y)\,dy = \left(n + \frac{1}{2}\right)\pi; \tag{11}$$

this well known (Sommerfeld-like) condition is the one that applied to the harmonic oscillator gives the discrete spectrum.

Let us now analyze how we can apply these techniques to our problem. The equation that we want to solve is Eq. 3.1.1–(1) that can be written into the convenient form

$$f_n''(x) + k_n^2(x) f_n(x) = 0, \tag{12}$$

where $k_n^2(x) \equiv \omega_n^2 - \lambda(x)$. And from now on let us restrict ourselves to *DD* boundary conditions. Moreover in order to obtain results we have to make a crucial assumption on λ, that is to say

$$\lim_{x\to 0} x^2 \lambda(x) = 0; \tag{13}$$

this actually is a requirement that $\lambda(x)$ does not have a bad behavior at the boundaries. Remember that the symmetry of the problem also requires that $\lambda(x) = \lambda(L - x)$.

This time the correct combination of solutions (6) is determined, as we will prove in a moment by the boundary conditions and by the property (13). Consider the following normalized combination of solutions in the classically allowed region:

$$f_n^W(x) = \frac{C_n}{\sqrt{k_n(x)}} \sin\Big(\int_{\epsilon_n}^{x} k_n(y)\,dy\Big), \tag{14}$$

where $C_n = \sqrt{2/\int_{\epsilon_n}^{L-\epsilon_n} dx/k_n(x)}$, and ϵ_n and $L-\epsilon_n$ (because of the already mentioned symmetry) are the classical turning points — the points where $k_n(x) = 0$. In general two things can happen: Either λ is finite at the boundaries, or it diverges (consistently with Eq. (13)). If it doesn't diverge, then it means that there exists a maximum value for λ, let us call

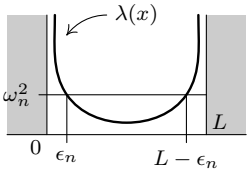


Fig. 2. Turning points.

it $\bar{\lambda}$. In this case it is clear that for n sufficiently big $\omega_n^2 > \bar{\lambda}$ (notice that this surely happens because of 3.1.1–(5)). But then it is clear that the turning points are $x = 0$ and $x = L$, that is to say $\epsilon_n = 0$ for all n that are big enough. Then Eq. (14) is consistent with *DD* boundary conditions if $f_n^W(L) = 0$ which implies

$$\int_0^L k_n(y)\, dy = m\pi, \quad m = 1, 2, \dots \,. \tag{15}$$

On the other hand if λ is not bounded, because of condition (13), and from the definition of the classical turning points, we have that $\lim_{\omega_n \to 0} \omega_n \epsilon_n = 0$, which means that ϵ_n becomes 0 way faster than the typical wavelength of the n-th eigenstate $1/\omega_n$. Then here we get exactly the same result that is written in Eq. (14), only that the Sommerfeld condition (15) must be replaced by the analogous

$$\int_{\epsilon_n}^{L-\epsilon_n} k_n(y)\, dy = m\pi, \quad m = 1, 2, \dots \,. \tag{16}$$

All this actually means that for large n the eigenfunctions and eigenvalues are well approximated by

$$f_n(x) \approx \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}, \qquad \omega_n^2 \approx \frac{n\pi}{L}. \tag{17}$$

With this result at hand we can now show that λ is actually divergent at $x = 0$ (and at $x = L$) as it is sketched in Fig. 2. (Actual plots of the function will be displayed later on.)

Behavior near the boundaries. Let us start from the behavior of $\tilde{\sigma}^2$ near the boundaries: Looking back at Eq. 3.1.2–(3) we realize that we just have to study the behavior of the sum $N \sum_{n \geq 1} [f_n^2/(2\omega_n) - 1/(2\omega_n L)]$. Notice that we can always split such sum into the sum on $n \leq \bar{n}$ and the sum on $n > \bar{n}$, where $\bar{n}$ is large enough so that for all $n > \bar{n}$, Eq. (17) is a good approximation.

The finite sum up to $\bar{n}$ contributes with just a constant to the behavior around zero, since $\lim_{x \to 0} f_n(x) = 0$, and therefore can be neglected, for the same reason we can add the sum from 1 to $\bar{n}$ of the eigenfunctions in Eq. (17):

$$\begin{aligned} \tilde{\sigma}^2(x) = -N \sum_{n \geq 1} \frac{1}{2\omega_n} \left(f_n^2(x) - \frac{1}{L} \right) &\simeq \frac{N}{2\pi} \sum_{n \geq 1} \frac{1}{n} \cos \frac{2n\pi x}{L} \\ &= -\frac{N}{2\pi} \log \left(2 \sin \frac{\pi x}{L} \right). \end{aligned} \tag{18}$$

This means that as $x \to 0$, $\sigma^2(x)$ (which has the same behavior of $\tilde{\sigma}^2$ apart from a constant) behaves as

$$\sigma^2(x) \simeq \frac{N}{2\pi} \log \frac{1}{x}. \tag{19}$$

The behavior of λ near the boundaries can be determined through the first gap equation $\sigma''(x) - \lambda(x)\sigma(x) = 0$; which immediately tells us that

$$\lambda(x) \simeq \frac{1}{2x^2 \log(1/x)}. \tag{20}$$

So this actually proves that both λ and σ are divergent near the boundaries. Notice that the behavior of λ displayed in Eq. (20) is compatible with our earlier assumption (13), this is in fact a consistency check.

Now one should ask what is the reason of such divergences: The physical reason. After a little thinking it becomes apparent that it must be so; far away from the boundaries the UV divergences are canceled by the renormalized size of $\mathbf{C}\mathrm{P}^{N-1}$, r, that — as can be seen from Eq. 3.1.2.1–(15) — is logarithmically divergent, just like in the two-dimensional case (this is because such divergences are local, so they do not see the boundary effects).

However near the boundaries the n_i fields are constrained, so the σ field must cancel the constant r. This is what happens for Dirichlet and Neumann conditions; periodic conditions, on the other hand, do not have such divergences (as we have already seen) since they do not make distinction between internal points and boundary points: Boundary points are special only with respect to the way in which compactification is done, but then they become exactly as any other point where the UV divergences are canceled by the renormalization of the coupling constant.

Just to complete this section on behavior near boundaries let us quickly mention what happens to the eigenfunctions $f_n(x)$ near $x = 0$ and $x = L$.

As it is shown in the appendix of the article by Konishi, Bolognesi and Ohashi [9], the behavior of the eigenfunctions at the boundary can be understood analytically expanding the f_ns and then using Eq. 3.1.1–(1) to find a recursion relation between the different coefficients of the expansion. We are then able to derive that the leading terms are

$$f_n(x) \simeq \frac{x}{\sqrt{\log(1/x)}} \quad \text{around } x = 0, \tag{21}$$

and

$$f_n(x) \simeq \frac{L - x}{\sqrt{\log(1/L - x)}} \quad \text{around } x = L. \tag{22}$$

An interesting point is that these asymptotic values not only do satisfy *DD* , but also *NN* boundary conditions.

Regularization of the gap equations. Before moving on to the numerical resolution of the gap equations, we have to spend some words on the issue of regularization of the gap equations.

By direct inspection of Eq. 3.1.2–(1) and 3.1.2–(3) we see that only the constant part of the second gap equation needs regularization because of the UV divergences that affect the sum $\sum_{n\geq 1} 1/\omega_n$, as one can easily understand using the WKB approximation on the higher modes.

One way to regularize such expression is to introduce a high-mode dumping factor $\exp(-\epsilon n)$, where ϵ, as usual, is a positive parameter. The sum over low energy modes has then a smooth limit as $\epsilon \to 0$, what we have to carefully study is the behavior of the sum over high modes as $\epsilon \to 0$. To be precise consider some index $\bar{n}$, so that for $n > \bar{n}$ the eigenfunctions f_n can be approximated by

the expression in Eq. (17), then we can use the approximation

$$\frac{\pi}{L}\sum_{n>\bar{n}}\frac{e^{-\epsilon\omega_n}}{\omega_n}\simeq\sum_{n>\bar{n}}\frac{e^{-(n\pi/L)\epsilon}}{n}, \tag{23}$$

and ask ourselves what happens to this expression as $\epsilon\to 0$. Now rewrite this expression adding and subtracting to this sum the first terms from $n=1$ to $n=\bar{n}$; the sum of the first $\bar{n}$ terms is convergent and one can immediately take the limit as ϵ goes to zero, and it gives the $\bar{n}$-th harmonic number $H_{\bar{n}}$ (remember definition 3.1.2.1–(6)): What is left is the $\sum_{n\geq 1}\exp[(n\pi/L)\epsilon]/n$. This sum is well known and is exactly

$$\sum_{n\geq 1}\frac{e^{-(n\pi/L)\epsilon}}{n}=-\log\left(1-e^{-\epsilon\pi/L}\right), \tag{24}$$

so we have just proved that

$$\frac{\pi}{L}\sum_{n>\bar{n}}\frac{e^{-\epsilon\omega_n}}{\omega_n}=-\log\left(1-e^{-\epsilon\pi/L}\right)-H_{\bar{n}}. \tag{25}$$

Moreover since at this point we can take $\bar{n}$ as big as we please, there is a nice approximation for the harmonic numbers that can be derived using Euler's summation formula (see Knuth *Fundamental Algorithms* [30]), we have

$$H_n=\log n+\gamma+\frac{1}{2n}-\frac{1}{12n^2}+\frac{1}{120n^4}+O\left(\frac{1}{n^6}\right), \tag{26}$$

where $\gamma=0.5772156649\ldots$ is Euler's constant: Here there are some values of H_n for small n up to $n=14$:

$$\begin{array}{rccccccccccccccc} n &= 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & \ldots \\ H_n &= 0 & 1 & \frac{3}{2} & \frac{11}{6} & \frac{25}{12} & \frac{137}{60} & \frac{49}{20} & \frac{363}{140} & \frac{761}{280} & \frac{7129}{2520} & \frac{7381}{2520} & \frac{83711}{27720} & \frac{86021}{27720} & \frac{1145993}{360360} & \frac{1171733}{360360} & \ldots \end{array} \tag{27}$$

for a more complete list, with harmonic numbers up to $n=30$, see Appendix A of Knuth book *Fundamental Algorithms*. Then using the approximation in Eq. (26), and expanding around $\epsilon=0$ we can rewrite Eq. (25) as

$$\frac{\pi}{L}\sum_{n>\bar{n}}\frac{e^{-\epsilon\omega_n}}{\omega_n}=-\log\frac{\pi\epsilon}{L}-\log\bar{n}-\gamma+O(\epsilon)+O(1/n^6). \tag{28}$$

Inserting this result in the constant part of the gap equation we obtain

$$\frac{N}{L}\sum_{n=1}^{\bar{n}}\frac{1}{2\omega_n}+\frac{1}{L}\int_0^L\sigma^2(x)\,dx+\frac{N}{2\pi}(-\log\epsilon-\log\mu(\bar{n})-\gamma)-r=0, \tag{29}$$

if we define $\mu(n)=n\pi/L$.

The crucial point here, as we have already pointed out in the previous section, is that UV divergences must be canceled out locally, so the correct value of r to be used in this formula has already been evaluated while analyzing

the translational invariant case, and it is exactly given by Eq. 3.1.2.1–(14). Substitution of this equation into Eq. (29) eventually gives

$$\frac{N}{L}\sum_{n=1}^{\bar{n}}\frac{1}{2\omega_n}+\frac{1}{L}\int_0^L \sigma^2(x)\,dx-\frac{N}{2\pi}\log\frac{2\mu(\bar{n})}{\Lambda_{\rm CP}}=0. \tag{30}$$

We are now ready to study the gap equations numerically.

3.1.2.3. Numerical Analysis. We will now outline and present the result of a numerical method to find the solution of the gap equations that has been used in [9].

Notice that the regularization used in Eq. 3.1.2.2–(30) is particularly suited for numerical analysis since the cutoff has been chosen to be on the number of modes — a parameter easily controllable inside a computer (it is an integer number).

Let us now outline the algorithm that we will use in order to compute $\lambda(x)$ and $\sigma(x)$ numerically. Start with an initial $\lambda(x)$ that is consistent with Eq. 3.1.2.2–(13), for example $\lambda\equiv 0$, then use this value of λ to evaluate the eigenvalues and eigenfunctions via the equation $-f_n''(x)+\lambda(x)f_n(x)=\omega_n^2 f_n(x)=0$. At this point we use Eq. 3.1.2.2–(30) together with the varying part of this gap equation to find $\sigma(x)$, invert the first gap equation $\sigma''(x)-\lambda(x)\sigma(x)=0$ to obtain the new $\lambda(x)$ and start all over again until the method converges to stable solutions. Of course as one might expect there are several subtleties that one must take into account, and we will try to address them in what follows.

The program that we are going to explain is written in the *Mathematica* programming language. For simplicity we do not bother writing the output routines. The codes start with the declaration of some variables:

```
L=10;                          (* this is the length of the model *)
Lambda=0.004;                          (* this is the mass scale *)
Nmax=100;                    (* cutoff in the numbers of modes *)
MAX=4000;                      (* maximum number of iterations *)
CEll=0.01;                          (* option of NDEigensystem *)
d=73;                                (* determines the spacing *)
                          (* of the points chosen to do the fit *)
a=0.0001;                  (* the solution is found in [a..L-a] *)
```

In order to have a cleaner code we will have a function $V[x\_]$ that represent $\lambda(x)$, and a function $S[x\_]$, that represent σ^2. We then initialize λ to zero:

```
V[x_]=0;                                      (* Initialize \lambda *)
```

Now as we anticipated the procedure that we use is a recursive procedure, so we will need a loop:

```
For[j=1, j<12, j++,
 {vals, funs}=NDEigensystem[{-Laplacian[u[x], {x}] + V[x] u[x],
                              DirichletCondition[u[x] == 0,
```

```
                              True]},
                             u[x], {x, a, L-a}, Nmax,
   Method->{``Eigensystem''->{``Arnoldi'',
                            ``MaxIterations''->MAX},
           ``SpatialDiscretization'' -> {``FiniteElement'',
                                         ``MeshOptions'' ->
                              {``MaxCellMeasure'' -> CEll}}}
                             ];
```

In order to find the eigenvalues and eigenfunctions here we have used the *Mathematica* subroutine **NDEigensystem**$(L(u(x,y,\ldots)),u,(x,y,\ldots)\in\Omega,n)$, that generates the first n eigenvalues and eigenfunctions of the problem $Lu=\omega u$ in the region Ω. Notice also that some options have been specified: Dirichlet boundary conditions and some details on the computational method used (for further details we refer to the Wolfram Documentation Center). The "for" cycle then continue to compute the solutions

```
S1[x_]=-((1/2) vals^(-1/2).funs^2 - (1/(2 L))
       vals^(-1/4).vals^(-1/4));
CC= -(1/(2 L)) vals^(-1/4).vals^(-1/4)
    +(1/(2 Pi)) Log[2 Pi(Nmax/L)]-(1/(2 Pi)) Log[Lambda];

f={};                                  (* initialize an array *)
For[n=1, n<d-1, n++
 AppendTo[f,{(L n)/d, S1[(L n)/d]+CC}]];
                (* the function that we use to do the fit is*)
model=-(1/(2Pi)) (Log[x]+Log[L-x]+C1+C2 Cosh[C3 (x-(L/2))]);
fit1=FindFit[f,model,{C1,C2,C3},x];  (* this finds C1,C2,C3 *)
firstFIT=Function[{x}, Evaluate[model/.fit1]];  (* function *)
```

Although now we already have a function that represent σ^2, namely $CC+S1$, we improve the result by fitting the difference between the values found by **NDEigensystem** and the values obtained from the fitted function *firstFit*.

```
ff={};              (* we initialize a new vector of points *)
For[n=1, n<d-1, n++
 AppendTo[ff,{(L n)/d, S1[(L n)/d]+CC-firstFIT[(L n)/d]}]];
          (* and we fill it with the difference S1-firstFit *)
model2=j1 Sin[x (Pi/L)]+j3 Sin[3 x (Pi/L)]+j5 Sin[5 x (Pi/L)];
fit2=FindFit[ff,model2,{j1,j3,j5},x];
secondFIT=Function[{x}, Evaluate[model2/.fit2]];
```

This finally gives us the function $\sigma^2(x)$. Notice that we could have added more terms to the expansion that defines *fit2*, however that wouldn't have done much difference. (You can try it!)

```
S[x_]:=firstFIT[x]+secondFIT[x];
```

Now for the delicate part of the program. As we have argued, we have to invert $\sigma'' - \lambda\sigma = 0$ to update λ and start the cycle again. If one performs this inversion naively , saying that

$$\lambda = \frac{\sigma''}{\sigma} \tag{1}$$

the program will converge only for small values of $L\Lambda$; this because for length larger than the scale of the problem, in the middle of the interval, σ becomes very small, so the division by σ in (1) will give accuracy issues. The solution that we adopt here to extend the domain of convergence of the program is to slow down the update of λ by taking a mean between the new value evaluated in the current "for" cycle with the previous value. That is:

```
 V1[x_]= (1/2) S[x]^(-1) S''[x] -(1/4) S[x]^(-2) (S'[x])^2;
 V2[x_]=V[x];                    (* this is the previous value*)
 V[x_]=(1/15) (V1[x]+14 V2[x]);        (* this is the update *)
]                                              (* end of for *)
```

The output of the above program, as you can see, are the functions $\lambda(x)$ and $\sigma^2(x)/N$: Results of running this program with different values of L and Λ are presented in the following pages, in Fig. 1(a) to Fig 1(d) and in Fig 2(a) to Fig 2(b).

Fig. 2(a) to Fig 2(f) could not have been obtained without the device explained just now to slow down the λ convergence.

Beware of bugs in the above code;
I have only proved it correct, not tried it.
— DONALD E. KNUTH (1977)

Discussion of the results. Let us now try to understand this results from a physical point of view.

Fig 1(a), 1(b) show plots for a fixed value of $L = 1$ and various values of $\Lambda_{\rm CP}$ ranging from 0.2 to 1; Fig. 1(c), 1(d) and Fig. 2(a) and 2(b) instead show plots for $\Lambda_{\rm CP} = 1$ and various values of L.

First of all notice that not all these results are independent; call $\lambda(x; L, \Lambda)$ and $\sigma(x, L, \Lambda)$ the results of the gap equations that one obtains setting the length of the interval equal to L and $\Lambda_{\rm CP} = \Lambda$: One can then show that results computed at the same $L\Lambda$ are related by simple relations due to dimensional arguments.

From the term $\lambda\sigma^2$ in the $\mathbf{C}\mathrm{P}^{N-1}$ Lagrangian one immediately sees that $[\lambda] = 2$, since $[\sigma] = 0$ in a two dimensional theory. From this, plus the fact that $L\Lambda$ is the only non trivial dimensionless quantity, if we perform a scale transformation $L \to L/\alpha$ and $\Lambda \to \alpha\Lambda$ that preserves $L\Lambda$ then we must have

$$\lambda(x; L, \Lambda) = \frac{1}{\alpha^2}\lambda(x/\alpha; L/\alpha, \alpha\Lambda), \qquad \sigma(x; L, \Lambda) = \sigma(x/\alpha; L/\alpha, \alpha\Lambda). \tag{2}$$

Now notice that from the plots in Fig. 1 even thought we start to see that both σ and λ assume a constant behavior in the middle of the interval it is not clear

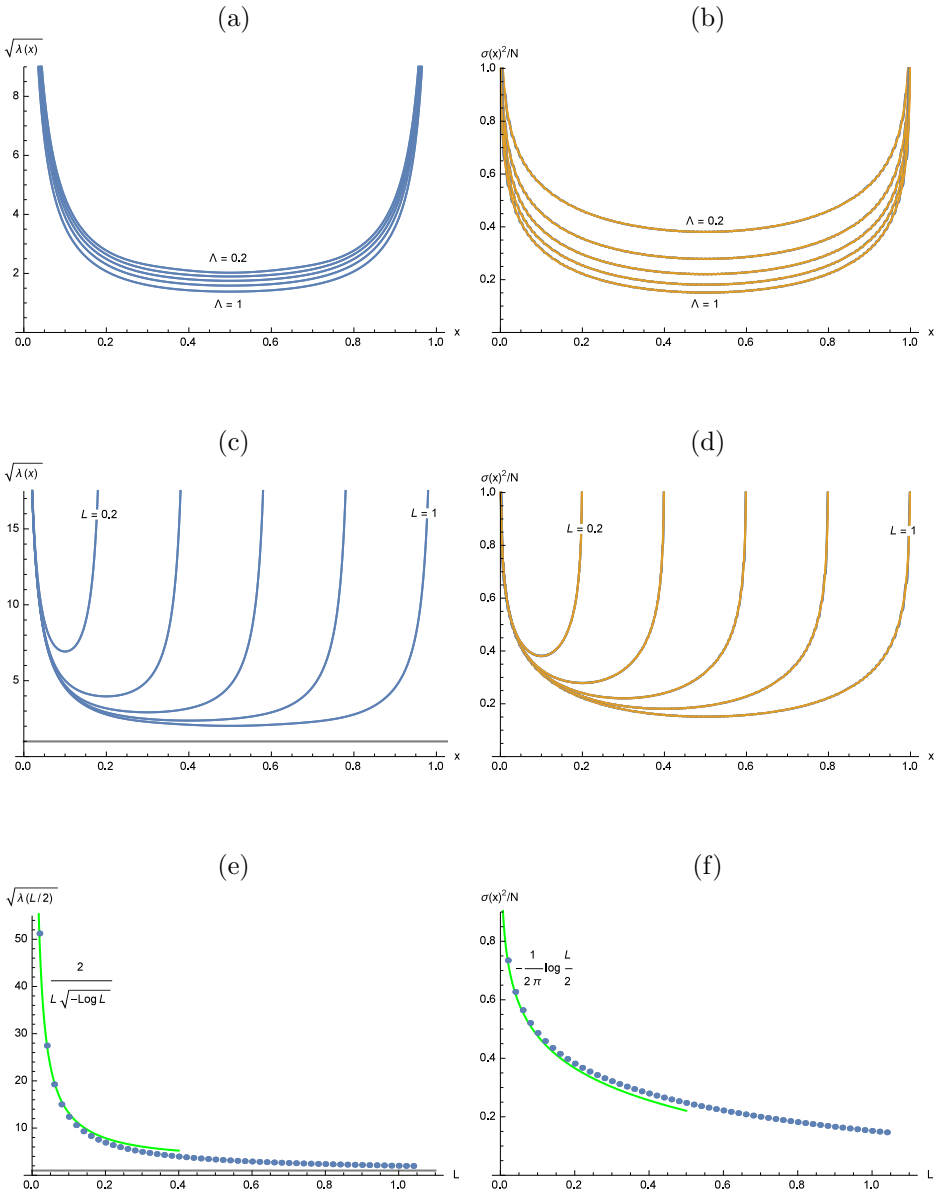


Fig. 1. Results of the program discussed in Section 3.1.2.3. Plots (a) and (b) are obtained keeping $L = 1$ and making Λ vary from 0.2 to 0.1. Plots (c) and (d), on the other hand, are the results if one keeps $\Lambda = 1$ and changes L from 0.2 to 1. The data in (e) and (f) show the dependence of the value of $\sqrt{\lambda(L/2)}$ and $\sigma^2(L/2)/N$ on L with $\Lambda = 1$.

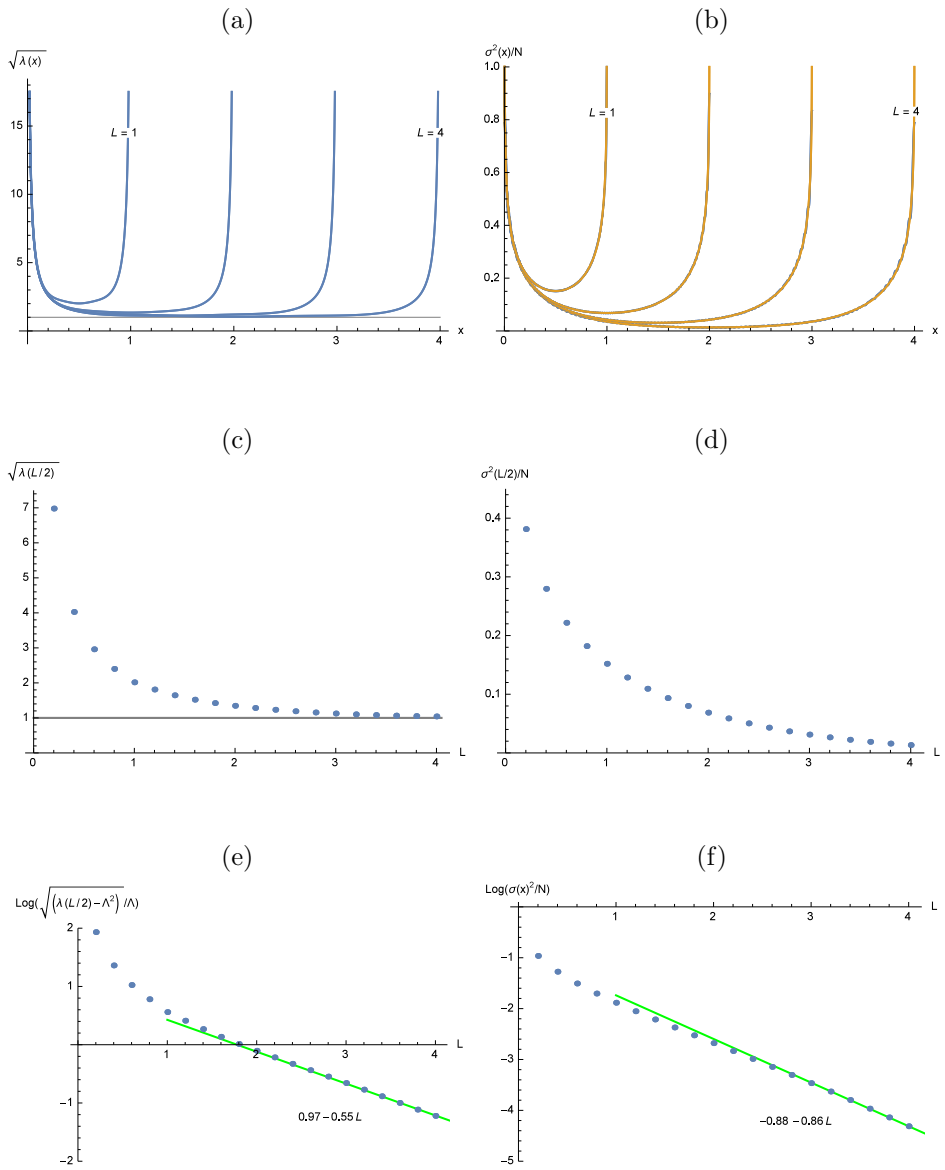


Fig. 2. Results of the program discussed in Section 3.1.2.3. Plots (a) and (b) are obtained keeping $\Lambda = 1$ and making L vary from 1 to 4. Plots (c) and (d) show the dependence of $\sqrt{\lambda(L/2)}$ and $\sigma^2(L/2)/N$ on L with $\Lambda = 1$; in figures (e) and (f) such dependence is fitted in a logarithmic plot with two lines.

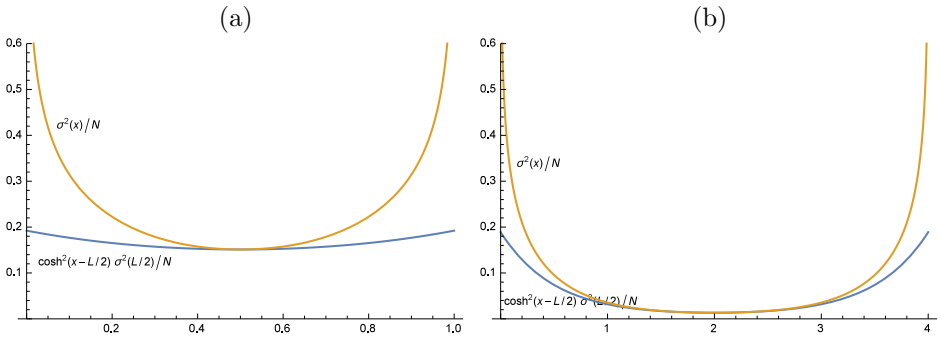


Fig. 3. Comparison between numerical data and solutions found assuming $\lambda = \Lambda^2$ for $\Lambda = 1$ and $L = 1$ and $L = 4$.

what the asymptotic value will be. In fact if we plot the values of σ and λ at $x = L/2$ for various values of L keeping Λ fixed to 1, as it is done in Fig. 1(e) and 1(f), in this range of values, we can just say that they seem to have a monotonic behavior but (especially for σ) we have no clue of what the limiting value would be.

This situation changes completely once we take into account the results of numerical simulations that we obtain for larger values of $L\Lambda$. So, considering now Fig. 2(a) and 2(b), from these pictures we can distinguish an expected pattern: For large L the values of λ and σ approach the values that they should have if the theory were defined on the plane, $\lambda \to \Lambda^2$ and $\sigma \to 0$.

This behavior is particularly clear for the point in the middle of the interval, as it is shown in Fig. 2(c) and 2(d), where the value of the solutions at $x = L/2$ are plotted against various values of L. One can actually try to fit those data assuming an exponential falloff; the result of such fit can be seen in Fig. 2(e) for λ and in Fig. 2(f) for σ in a logarithmic plot. The numerical results that has been found in [9] is indeed

$$\lambda(L/2) = \Lambda^2 + \Lambda^2 e^{0.97-0.55L}, \qquad \sigma^2(L/2) = N e^{-0.88-0.86L}. \tag{3}$$

Another indication of the fact that the fields λ and σ rapidly approach to the value they assume at infinite L might be the following: Suppose that we assume that around $x = L/2$ the field λ is Λ^2 if L is big enough, and let us try to use the gap equation $\sigma'' - \lambda\sigma = 0$ to derive the behavior of σ around this point. This equation becomes $\sigma'' - \Lambda^2\sigma = 0$ that has as solutions $\exp(\pm\Lambda x)$. However because of the symmetry of the problem $\sigma(x) = \sigma(L - x)$ the solution is fixed to be

$$\sigma(x) = \sigma(L/2)\,\mathrm{Ch}(x - L/2). \tag{4}$$

In Fig. 3(a) and 3(b) such function is plotted and compared to the numerical results we previously showed for $\Lambda = 1$ and $L = 1$ (in Fig. 3(a)) and $L = 4$ (in Fig. 3(b)); as one can see for $L = 4$ we have a large region in the middle of the interval in which the two graph cannot be distinguished.

The energy has a large number of different forms,
and there is a formula for each one.
— RICHARD P. FEYNMAN, The Feynman Lectures on Physics (1963)

3.2. THE ENERGY

IN THIS SECTION we address the problem of the energy in finite size $\mathbf{C}\mathrm{P}^{N-1}$ theories. This problem is crucial in view of a good comprehension of the confinement mechanism that dualizes the Abrikosov effect for superconductors since this energy corresponds to the "orientational" energy of the string.

The energy, and its dependence on the size of the strip on which $\mathbf{C}\mathrm{P}^{N-1}$ is defined can be analyzed in details under the assumption of translational invariance (that, as we discussed, is satisfied when we assign periodic conditions).

The same analysis for general boundary conditions is much more difficult, since even the proof that the energy density can be renormalized is highly nontrivial. This is because unlike the translational invariance case we cannot express the propagator, and therefore the gap equations in a closed form, so that in order to have analytical results we need to rely on approximations (such as the WKB approximation) that holds in specific regimes, with the result that we need to be extremely careful to correctly handle all the limiting processes that are involved in the calculations.

3.2.1. Energy for the system on a cylinder

Let us continue the discussion we begun in Section 3.1.2.1; in that section we proved that the periodic boundary conditions are consistent with the translational invariant solutions of gap equations.

However we can say much more in this case — as it has been shown by Shifman [40] — since we can actually derive the total energy and the dependence of this from the length of the interval (or if you want from the length of the circumference that defines the cylinder shaped region that constitutes the domain of our model).

As it is argued in Shifman in the case of a compactified strip we can assume that the vacuum solutions have a constant non-vanishing gauge field A_1. Then the propagator and the gap equation will depend on this additional parameter. The defining equation for the propagator will have extra pieces that come from the fact that the differential operator that is now relevant is $(i\partial + A)^2$, then Eq. 3.1.2.1–(2) now becomes

$$(-\partial^2 + 2iA_1\partial_x + A_1^2 + m_2)\,\mathscr{D}(x,t;y,s) = \delta(t-s)\delta(x-y). \tag{1}$$

The solution of such equation, as one can easily verify, is

$$\mathscr{D}(x,y;t,s) = \frac{1}{2\pi}K_0(m\sqrt{(x-y)^2+(t-s)^2})e^{A_1\cdot(x-y)}, \tag{2}$$

so that the full solution that takes care of all mirror images is given as usual by the superposition principle

$$\mathscr{D}(x,y;t,s) = \frac{1}{2\pi}\sum_k K_0(m\sqrt{(x-y+kL)^2+(t-s)^2})e^{A_1\cdot(x-y+kL)}. \qquad (3)$$

Now that we have the propagator, we can derive the energy using the following formula that will be derived carefully in the next section (see in particular the subsection about the "propagator representation of the energy")

$$E = \int_0^L \left\{ N\lim_{\epsilon\to 0}\left(2\frac{\partial^2}{\partial\epsilon^2}+\frac{1}{2}\frac{\partial^2}{\partial x^2}\right)\mathscr{D}_\epsilon(x) + \sigma'^2 + \lambda(\sigma^2-r)\right\} dx. \qquad (4)$$

Now many simplifications occur in such formula. First of all in this case $\mathscr{D}_\epsilon$ does not depend on position, so the derivatives with respect to x are zero. Moreover since both $\lambda \equiv m^2$ and σ are constant the integral just gives a multiplicative factor L. Then the σ gap equation in the translational invariant case is $\lambda\sigma = 0$, and we choose $\sigma = 0$ (what in literature is called "confinement phase"). Then using Eq. 3.1.2.1–(14), and dropping the divergent terms $1/\epsilon^2$ we find

$$E(L) = \frac{Nm^2L}{2\pi}\left(\frac{1}{2}-\log\frac{m}{\Lambda_{\rm CP}}\right) - \frac{2Nm^2L}{\pi}\sum_{k\geq 1}\frac{K_1(kmL)}{kmL}\cos(kA_1L). \qquad (5)$$

Since we have now an explicit expression for the energy, the gap equation is simply given by the derivative of Eq. (5) with respect to m^2, and A_1 is fixed to zero by the requirement that $\partial E/\partial A_1 = 0$.

Then finally we can study the dependence of the energy on the length of the interval in the limit $L \gg 1/\Lambda_{\rm CP}$. In this limit Bessel functions behave like an exponential, and we can use the leading order approximation $m = \Lambda_{\rm CP}$.

$$E(L) \approx \frac{N\Lambda_{\rm CP}^2}{4\pi}L - N\sqrt{\frac{2\Lambda_{\rm CP}}{\pi L}}e^{-\Lambda_{\rm CP}L} + \dots . \qquad (6)$$

This is the behavior of the energy at large L of the $\mathbf{C}\mathrm{P}^{N-1}$ model defined on a cylinder (that is on a strip with periodic boundary conditions).

I move that the meeting adjourn,
for the immediate adoption of more energetic remedies.
— LEWIS CARROLL, Alice's Adventures in Wonderland (1865)

3.2.2. Energy under general boundary conditions

Our task now is to study the energy of $\mathbf{C}\mathrm{P}^{N-1}$ on a finite size strip in the large N limit with *DD* or *NN* boundary conditions.

So the first thing we should do is to derive a convincing expression for the energy/energy density, starting from the Lagrangian of the model in Eq. 3.1–(4).

In order to do this we will follow the standard procedure prescribed by QFT (see [53]): Compute the Hamiltonian density and then take the vacuum

expectation value to derive the required expression for the energy density in the large N limit.

In order to derive the Hamiltonian density it is easier to work with the expression of the Lagrangian that still contains the n_i fields with $A_\mu = 0$. First of all consider the part of the Lagrangian that contains σ: $\mathscr{L}_\sigma = (\partial_\mu\sigma\partial^\mu\sigma - \lambda(\sigma^2 - r))[x \in S]$; for this field it is safe to replace the dependence $\sigma(x,t)$ with $\sigma(x)$, $x \in [0\,..\,L]$. Then the energy density of this part is just $\langle\mathscr{H}_\sigma\rangle = \langle(\sigma'^2 + \lambda(\sigma^2 - r))[0 \le x \le L]\rangle = (\sigma'^2 + \lambda(\sigma^2 - r))[0 \le x \le L]$.

The part of the Lagrangian that depends on n is a little bit trickier to handle, so let us do all the steps required carefully. The Lagrangian that we are talking about is

$$\mathscr{L}_n = \Big(\sum_{j=2}^{N}[\partial_\mu n_j^*\partial^\mu n_j - \lambda n_i^* n_i]\Big)[0 \le x \le L]; \tag{1}$$

taking the Legendre transform, the Hamiltonian density turns out to be

$$\mathscr{H}_n = \Big(\sum_{j=2}^{N}[\partial_t n_j^*\partial_t n_j + \partial_x n_j^*\partial_x n_j + \lambda n_i^* n_i]\Big)[0 \le x \le L]. \tag{2}$$

This time in order to find the energy density we need to compute three expectation values: $\langle\partial_t n_i^*(x,t)\partial_t n_i(x,t)\rangle$, $\langle\partial_x n_i^*(x,t)\partial_x n_i(x,t)\rangle$ and $\langle n_i^*(x,t)n_i(x,t)\rangle$.

The last one we have written is the easiest, since it is just the propagator of the n_i fields, for the other two instead we need to perform an explicit calculation. We have

$$\langle\partial_x n_i^*(x,t)\partial_x n_i(x,t)\rangle = \lim_{x\to y}\partial_x\partial_y\langle n_i^*(x,t)n_i(y,t)\rangle = \sum_{n\ge1}\frac{f'^2(x)}{2\omega_n}, \tag{3}$$

and for the term with time derivatives:

$$\begin{aligned}\langle\partial_t n_i^*(x,t)\partial_t n_i(x,t)\rangle &= \lim_{t\to s}\partial_t\partial_s\langle n_i^*(x,t)n_i(x,s)\rangle \\ &= \lim_{t\to s}\partial_s\sum_{n\ge1}(-i)\omega_n\frac{e^{-i\omega_n|t-s|}}{2\omega_n}([t>s]-[t<s])f_n^2(x) \\ &= \lim_{t\to s}\sum_{n\ge1}\Big(\frac{\omega_n}{2} - i\,[t=s]\Big)\,f_n^2(x)e^{-i\omega_n|t-s|},\end{aligned} \tag{4}$$

where we took advantage of the Iverson's notation to write the well known fact that the derivative of the sign function $[t>s]-[t<s]$ is twice the delta function $[t=s]$.

As one can see from the last expression in Eq (4) the limit is divergent because of the delta function. This divergence is just a constant, even if it is infinite, and therefore can be dropped. (Actually the same calculation can be done using standard canonical quantization and if one orders the operators carefully obtains the same result as here but without the divergent term.)

The energy density for the n fields is therefore (remember that, since we work in the large N approximation we always replace $N-1$ by N)

$$\mathscr{E}_n(x) = N\Big\{\sum_{n\geq 1}\left(\frac{\omega_n}{2}f_n^2(x) + \frac{{f_n'}^2(x)}{2\omega_n} + \lambda\frac{f_n^2}{2\omega_n}\right)\Big\}[0\leq x\leq L], \tag{5}$$

and by noticing that $(f_n^2(x))'' = 2f_n'(x)f_n(x) + 2f_n(x)f_n''(x)$, and using Eq. 3.1.1–(1) we obtain an equivalent form for the energy, namely:

$$\mathscr{E}_n(x) = N\Big\{\sum_{n\geq 1}\left(\omega_n + \frac{1}{4\omega_n}\frac{d^2}{dx^2}\right)f_n^2(x)\Big\}[0\leq x\leq L]. \tag{6}$$

The total energy density is simply the sum of the energy density for the σ field and the one we have just derived: $\mathscr{E}(x) = \mathscr{E}_n(x) + \mathscr{E}_\sigma(x)$. Let us write the whole expression for example using Eq. (5)

$$\mathscr{E}(x) = \Big\{N\sum_{n\geq 1}\left(\frac{\omega_n}{2}f_n^2(x) + \frac{{f_n'}^2(x)}{2\omega_n} + \lambda\frac{f_n^2}{2\omega_n}\right) + \sigma'^2 + \lambda(\sigma^2 - r)\Big\}[0\leq x\leq L]. \tag{7}$$

We know that r has a divergent behavior as shown in Eq. 3.1.2.1–(14), in fact it has a very bad logarithmic divergence; then in order to have a renormalized gap equation we have to ask ourselves if the sum that came from the energy of the n_i fields can cancel this divergence or not. What we are asking then is whether the energy density can be made finite or not.

This question is not so straightforward to answer, it requires some thinking and manipulation, so we will dedicate to the problem the next section.

Before diving into the analysis of energy density (and eventually of energy) finiteness, let us mention a way to represent $\mathscr{E}_n$ in terms of the propagator, what we can refer to as the *propagator representation of the energy*.

Propagator representation of the Energy. As we have already done for the gap equation it is also possible to write the energy using a propagator representation in terms of $\mathscr{D}_\epsilon$.

Inspection of Eq. (6), and the comparison with the propagator in Eq. 3.1.2–(34) immediately tells us that

$$\mathscr{E}_n(x) = N\lim_{\epsilon\to 0}\left(2\frac{\partial^2}{\partial\epsilon^2} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\right)\mathscr{D}_\epsilon(x) \tag{8}$$

With this expression for the energy, it is clear that the divergence of the energy density can be in principle understood from the divergences of the propagator as $\epsilon\to 0$, in fact it is possible to try to find an expansion of the propagator.

Let us see the idea of how this expansion can be done in principle (this method has been suggested by Ohashi).

The starting point is the identity

$$\int_0^L \frac{\partial^2}{\partial\epsilon^2}\mathscr{D}_\epsilon(x)\,dx = \frac{1}{2}\sum_n e^{-\epsilon\omega_n}\omega_n; \tag{9}$$

then take the functional derivative of this equation with respect to $\lambda(x)$. Since we already know from Eq. 3.1.1–(13) that $\delta\omega_n = \int_0^L \delta\lambda(x) f_n(x)/(2\omega_n)\,dx$, then we get

$$\frac{\delta}{\delta\lambda(x)} \int_0^L \frac{\partial^2}{\partial\epsilon^2}\,\mathscr{D}_\epsilon(t)\,dt = \frac{1}{2}\frac{\partial}{\partial\epsilon}\big(\epsilon\,\mathscr{D}_\epsilon(x)\big). \tag{10}$$

The idea now is to take an expansion of $\mathscr{D}_\epsilon(x)$ in powers of ϵ and $\log\epsilon$ and obtain constraints on the coefficients of that expansion from this equation.

We already know that as $\epsilon \to 0$ the leading term of $\mathscr{D}_\epsilon$ is $1/(2\pi)\log 1/\epsilon$ since $N\,\mathscr{D}_\epsilon - r_\epsilon$ has to be finite (because $\sigma^2(x)$ is finite at finite x). And we also have an idea of why the expansion of $\mathscr{D}_\epsilon$ needs also to take care of $\log\epsilon$ terms since we know that in the $L \to \infty$ case the propagator is $(1/2\pi)K_0(m\epsilon)$ and with the known expansion 3.1.2.1–(5), the propagator expansion $\sum_{n\geq 0} d_n(x,\epsilon)\epsilon^n$ has

$$d_n(x,\epsilon) = \frac{1}{2\pi((n/2)!)^2}\left(-\log\frac{m\epsilon}{2} - \gamma + H_{n/2}\right)\left(\frac{m^2}{4}\right)^{n/2} \quad [n \text{ even}]. \tag{11}$$

Now, in general, substituting the expansion $\mathscr{D}_\epsilon = \sum_{n\geq 0} d_n\epsilon^n$ in Eq. (10) will give a recursive relation between the d_n coefficients, and in turn if we expand d_n in powers of $\log\epsilon$ — under the assumption that the d_ns do not contain essential singularities with respect to the variable $\log\epsilon$ — the same equation will give recursive relations for the coefficients of the expansion of the d_ns.

This method is, however, far too complicated to be handled and to be of any usefulness in the comprehension of the behavior of the energy. And even if it gives an expansion for small ϵ of the propagator, that indeed proves the finiteness of the energy density, it does so in terms of an unknown functional that prevents us from having a nice explicit formula for the energy density.

Because of all these reasons it will be more convenient to adopt another approach to attack the problem that we will now show.

3.2.2.1. Finiteness of energy density. We start by regularizing the energy density that we have derived in Eq. 3.2.2–(7).

This is simply done by multiplying the sum by the usual dumping factor $\exp(-\epsilon\omega_n)$, with ϵ positive and small. Moreover this regularization is consistent with the expression that we already have of the size of $\mathbf{C}\mathrm{P}^{N-1}$ expressed in Eq. 3.1.2.1–(14).

From now on we also drop the factor $[0 \leq x \leq L]$, in order to have slimmer expressions; with the understanding, if you want, that now the total energy is defined by the integral of the energy density on $[0\,..\,L]$, and not on the entire real line.

With this prescriptions Eq. 3.2.2–(7) becomes

$$\mathscr{E}(x) = N\sum_{n\geq 1}\left(\frac{\omega_n}{2}f_n^2(x) + \frac{{f_n'}^2(x)}{2\omega_n} + \lambda\frac{f_n^2}{2\omega_n}\right)e^{-\epsilon\omega_n} + \sigma'^2 + \lambda(\sigma^2 - r), \tag{1}$$

what we have to do then is to study the sums in the previous expression.

The key observation here is the fact that at x fixed the only divergences that can arise come from the infinite sum; in fact if we truncate the sum in Eq. (1) at some finite n the limit $\epsilon \to 0$ can be done without problems. On the other hand we already have studied in Section 3.1.2.2 a method to treat high modes: The WKB approximation.

Then for the purpose of studying the high-mode divergences, we can replace the eigenfunctions f_n and ω_n with the form in Eq. 3.1.2.2–(17). We don't even bother to start summing from an large n since the only difference that will make for a fixed x in $[0\,..\,L]$ is an additive constant (that does not depend on ϵ).

So with $f_n = \sqrt{2/L}\sin(n\pi x/L)$ and $\omega_n = n\pi/L$, the first two terms of the sum in Eq. (1) become

$$\frac{N}{L}\sum_{n\geq 1}\frac{n\pi}{L}e^{-\epsilon n\pi/L} = -\frac{N}{L}\frac{d}{d\epsilon}\sum_{n\geq 1}\left(e^{-\epsilon\pi/L}\right)^n = \frac{N\pi}{L^2}\frac{e^{\epsilon\pi/L}}{(1-e^{\epsilon\pi/L})^2}, \tag{2}$$

that expanded around $\epsilon = 0$ becomes

$$\frac{N}{\pi\epsilon^2} - \frac{N\pi}{12L^2} + \frac{N\pi^3\epsilon^2}{240L^4} + O(\epsilon^3). \tag{3}$$

The first two terms then give only an harmless divergence $1/\epsilon^2$, that can be easily subtracted. What about the term proportional to λ in Eq. (1)? It is immediately clear that this term will give a logarithmic divergence since ω_n in this approximation grows linearly. In fact direct calculation gives

$$\frac{N\lambda}{\pi}\sum_{n\geq 1}\frac{1}{n}\sin^2\frac{n\pi x}{L}e^{-\epsilon n\pi/L} = -\lambda\frac{N}{2\pi}\log\epsilon + \lambda\frac{N}{2\pi}\log\frac{2\sin(\pi x/L)}{\pi/L} + O(\epsilon^2), \tag{4}$$

that in fact is logarithmically divergent. This is a bad divergence that cannot be subtracted. However, as we have remarked this time we are lucky enough to have another term that diverges logarithmically, namely $-\lambda r$; moreover this other divergent part has the same coefficient and opposite sign, so it cancels exactly this divergent term. This actually proves that the energy density can be made finite!

There is yet another way to check that the cancellation involving the logarithmic divergence actually takes place, this method does not even use the WKB approximation and it is as easy as using the commutative and distributive law for real numbers: Just collect all terms proportional to λ together in Eq. (1)

$$\mathscr{E}(x) = N\sum_{n\geq 1}\left(\frac{\omega_n}{2}f_n^2(x) + \frac{f_n'^2(x)}{2\omega_n}\right)e^{-\epsilon\omega_n} + \sigma'^2 + \lambda\left(\sum_{n\geq 1}\frac{e^{-\epsilon\omega_n}}{2\omega_n}f_n^2 + \sigma^2 - r\right). \tag{5}$$

As one can see the terms proportional to λ reproduce the regularized gap equation, therefore all that part is exactly zero. With this argument we obtain also another even simpler expression for the energy density "on shell"

$$\mathscr{E}(x) = N\sum_{n\geq 1}\left(\frac{\omega_n}{2}f_n^2(x) + \frac{f_n'^2(x)}{2\omega_n}\right)e^{-\epsilon\omega_n} + \sigma'^2. \tag{6}$$

This expression, in principle, enable us to compute the energy density numerically.

The arguments we gave in the above discussion only prove the fact that the energy density can be made finite (at finite x), however they do not give the correct behaviors of the energy density. This is because the WKB approximation we have used, the one displayed in Eq. 3.1.2.2–(17), does not consider the dynamically generated mass m. This has as a consequence the appearance of the Lüscer term $-N\pi/12L^2$ that shouldn't be there because, as it is known, such term is associated with the number of massless degrees of freedom of the theory (for a discussion of this see the introduction to the paper [40]).

An apparent Paradox. In order to understand the subtleties of the problem at hand, it is instructive to discuss an apparent paradox and its solution that arises in this discussion of the energy.

Suppose that instead of Eq. 3.2.2–(5) we have used for the energy density of the n fields Eq. 3.2.2–(6). If one goes trough all the calculations similar to those we have just done, he will discover a very strange result: Even thought the divergence that goes like ϵ^2 is still there the logarithmic divergence disappear. This means that the nice cancellation with the term λr cannot take place any more. How is that possible? After all this two expressions are exactly equivalent!

The answer to this paradox is that we have to be careful with the WKB approximation. While in the Eq. 3.2.2–(5) the λ-dependence is correctly established by the explicit multiplicative factor, in the other expression, in order to have the right x dependence we have to consider the factor $(\omega_n^2 - \lambda)^{-1/4}$ from the WKB approximation (see Eq. 3.1.2.2–(6)).

This means that the correct eigenfunctions to use are

$$f_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \left(1 + \frac{\lambda}{2\omega_n^2} + \cdots\right). \tag{7}$$

Taking into account such corrections the correct results that we have established in the previous section are reproduced as expected.

3.2.2.2. Gap Equations from the Energy. In the previous sections we have derived various equivalent expressions for the energy. All those formulae for the energy density were essentially derived from a standard Hamiltonian approach. We worked with the initial n_i degrees of freedom and we implemented the large N expansion using the results we derived in Section 3.1.

In this section we consolidate the results we have obtained showing that when minimized the expressions we have derived actually give — as they should — the gap equations 3.1.2–(1). The starting point is Eq. 3.2.2.1–(5). As we will now show, the gap equation for σ is trivial to derive while that for λ requires some work.

First of all define the total energy of the system as

$$E[\lambda, \sigma] = \int_0^L \mathscr{E}(x)\, dx, \tag{1}$$

in this section we thought the total energy as a functional of λ and σ. The first gap equation $\delta E/\delta\sigma = 0$ can be immediately read off from Eq. 3.2.2.1–(5); after an integration by parts of the term σ'^2 we get:

$$\frac{\delta E}{\delta\sigma}[\lambda,\sigma] = -\sigma''(x) + \lambda(x)\sigma(x) = 0. \tag{2}$$

In order to derive the other equation, we must take the variation of $E[\lambda,\sigma]$ with respect to λ properly.

A little observation will simplify the discussion: The last term in Eq. 3.2.2.1–(5) is proportional to the wanted gap equation through the function λ, so that the variation of that term alone with respect to λ will give the correct answer. The problem then reduces to proving that the variation of the rest of the equation vanishes.

Let us call E_{12} the integral of the first two terms of the energy density: $E_{12} \equiv \int_0^L \{N\sum_{n\geq 1}[(\omega_n/2)f_n^2 + f_n'^2/(2\omega_n)]\exp(-\epsilon\omega_n) + \sigma'^2\}$. Then we have to prove — as we have argued — that $\delta E_{12} = 0$, where the variation of E_{12} is the amount of change in E_{12} when we change λ. The variation of such term then reads

$$\begin{aligned}\delta E_{12} =& \frac{N}{2}\sum_{n=1}^{\infty}\int_0^L \left[f_n^2(y)\delta\omega_n + 2\omega_n f_n(y)\delta f_n(y) - \frac{f_n'^2(y)}{\omega_n^2}\delta\omega_n + 2\frac{f_n'(y)}{\omega_n}\delta f_n'(y)\right] dy \\ &+ \frac{N}{2}\sum_{n=1}^{\infty}\int_0^L \lambda(y)\left[-\frac{f_n^2(y)}{\omega_n^2}\delta\omega_n + 2\frac{f_n(y)}{\omega_n}\delta f_n(y)\right] dy.\end{aligned} \tag{3}$$

We simplify this expression using the fact that $\lambda f_n = f_n'' + \omega_n^2 f_n$, obtaining:

$$\begin{aligned}\delta E_{12} =& \frac{N}{2}\sum_{n=1}^{\infty}\int_0^L \left[f_n^2(y)\delta\omega_n + 2\omega_n f_n(y)\delta f_n(y) - \frac{f_n'^2(y)}{\omega_n^2}\delta\omega_n + 2\frac{f_n'(y)}{\omega_n}\delta f_n'(y)\right. \\ &\left. - \frac{f_n''(y)f_n(y)}{\omega_n^2}\delta\omega_n - f_n^2(y)\delta\omega_n + 2\frac{f_n''(y)}{\omega_n}\delta f_n(y) + 2\omega_n f_n(y)\delta f_n(y)\right] dy,\end{aligned} \tag{4}$$

now observe that (after some eventual integration by parts) the first term in square brackets cancels with the sixth, the third cancels out with the fifth, the forth with the seventh.

Then the only terms left add up to $2N\sum_{n=1}^{\infty}\omega_n\int_0^L f_n(y)\delta f_n(y)\,dy$. However we have already noticed in Section 3.1.1 that because of normalization of the f_ns we have that $(f_n, \delta f_n) = 0$; this finally proves that $\delta E_{12} = 0$.

This is indeed an important consistency check, since not only it tests the form of the energy density (that at least now we know gives the correct gap equations) but also a bunch of other relations between the eigenfunctions f_n and the eigenvalues ω_n that we derived in the previous sections.

3.2.2.3. Energy and energy density at large L. Now that we have some ideas of how the energy density is made, we will show that we can formally prove that the energy density is constant in the interval $(0\,..\,L)$.

This is an highly nontrivial result since if we simply look at the expression in Eq. (6), and we try to understand the behavior of such expression near the boundaries we see that $\sigma'^2(x)$ is divergent. In fact from Eq. 3.1.2.2–(19) we know that at the boundaries $\sigma(x) \approx \sqrt{\log(1/x)}$, that means that $\sigma'^2(x) \approx 1/(x^2 \log(1/x))$ that is not only divergent but also has a non integrable behavior around $x = 0$. Saying that the energy density is constant means then that the sum in Eq. (6) exactly balances this bad behavior in such a way that the energy near the boundary assumes the same value (say $\mathscr{E}(L/2)$). Let us see how this result can be derived. The proof is actually quite simple: Start from Eq. (6) and take the derivative of this expression. This gives

$$\mathscr{E}'(x) = N \sum_{n\geq 1} \left(\omega_n f_n(x) f_n'(x) + \frac{f_n(x) f_n'(x)}{\omega_n} \right) + 2\sigma'(x)\sigma''(x), \tag{1}$$

then using the defining Eq. 3.1.1–(1) $f_n''(x) = [\lambda(x) - \omega_n^2] f_n(x)$, and the first gap equation $\sigma''(x) - \lambda(x)\sigma(x) = 0$ we obtain

$$\mathscr{E}'(x) = \lambda(x) \left(N \sum_{n\geq 1} \frac{f_n'(x) f_n(x)}{\omega_n} + 2\sigma'\,(x)\sigma(x) \right). \tag{2}$$

The term in parenthesis in nothing less that the derivative of the second gap equation $\sum_n f_n^2/(2\omega_n) + \sigma'^2 - r = 0$, then this immediately gives us the wanted result $\mathscr{E}'(x) = 0$.

With this result we can immediately tell that the total energy is

$$E = L\,\mathscr{E} \tag{3}$$

since $\mathscr{E}$ is a constant in $(0\,..\,L)$.

Moreover it enables us to study the behavior of the energy density at large L since we are now allowed to study such limit just for a fixed point. We choose to study the energy at the middle point because at that point the expression of the energy density further simplifies since because σ is symmetric with respect to that point we have that $\sigma'(L/2) = 0$.

Moreover for sufficiently large L, and at $x = L/2$ (as we have discussed at the end of Section 3.2.1.3) we can simply take $\lambda = m^2 = \Lambda_{\mathrm{CP}}^2$. In this case — assuming *DD* boundary condition for definiteness — the eigenvalues and eigenfunctions (and their derivatives) are known:

$$f_n(L/2) = \sqrt{\frac{2}{L}} \sin\frac{n\pi}{2}, \quad f_n'(L/2) = \sqrt{\frac{2}{L}} \frac{n\pi}{L} \cos\frac{n\pi}{2}, \quad \omega_n = \sqrt{\left(\frac{n\pi}{L}\right)^2 + \Lambda_{\mathrm{CP}}^2}. \tag{4}$$

In this approximation the energy density evaluated at the midpoint of the interval can be written as

$$\begin{aligned}\mathscr{E}(L/2) =& N\sum_{n\geq 1}\frac{1}{L}\Bigg(\sqrt{\Big(\frac{n\pi}{L}\Big)^2+\Lambda_{\rm CP}^2}\,\sin^2\frac{n\pi}{2}\\ &+\frac{(n\pi/L)^2}{\sqrt{\big(\frac{n\pi}{L}\big)^2+\Lambda_{\rm CP}^2}}\cos^2\frac{n\pi}{2}\Bigg)\exp\Bigg(-\epsilon\sqrt{\Big(\frac{n\pi}{L}\Big)^2+\Lambda_{\rm CP}^2}\Bigg);\end{aligned} \tag{5}$$

now for later convenience we can split the regularizing factor into the product of $\exp(-\epsilon n\pi/L)$ and $\exp[-\epsilon(\sqrt{(n\pi/L)^2+\Lambda_{\rm CP}^2}-n\pi/L)]$.

In order to get the behavior at large L the idea is to approximate the sum with an integral, this can be done by the formal replacing $n\pi/L\to z$ and $\sum_{n\geq 1}\pi/L\to\int_0^\infty dz$ then Eq. (5) becomes

$$\begin{aligned}\mathscr{E}(L/2)\simeq\frac{N}{\pi}\int_0^\infty dz\,\Bigg(&\sqrt{z^2+\Lambda_{\rm CP}^2}\,\sin^2\frac{zL}{2}\\ &+\frac{z^2}{\sqrt{z^2+\Lambda_{\rm CP}^2}}\cos^2\frac{zL}{2}\Bigg)e^{-\epsilon z}\exp\Big[-\epsilon\big(\sqrt{z^2+\Lambda_{\rm CP}^2}-z\big)\Big].\end{aligned} \tag{6}$$

Now since L is large the sinus and cosine terms give just a factor $1/2$, moreover one can change the integration variable sending $z\to\Lambda_{\rm CP}z$ and expand the last factor in Eq. (6) so that the expression further simplifies and becomes

$$\begin{aligned}\mathscr{E}(L/2)\simeq\frac{N\Lambda_{\rm CP}^2}{2\pi}\int_0^\infty dz\,\frac{2z^2+1}{\sqrt{z^2+1}}e^{-\epsilon\Lambda_{\rm CP}z}\Big[1-&\epsilon\Lambda_{\rm CP}(\sqrt{z^2+1}-z)\\ &+\epsilon^2\Lambda_{\rm CP}^2(\sqrt{z^2+1}-z)^2+\ldots\Big].\end{aligned} \tag{7}$$

The terms between square brackets in Eq. (7) that are at least of order ϵ^2, once one performs the integral, give to the energy density corrections that are vanishing as $\epsilon\to 0$; for example the integral of the term proportional to ϵ^2 behaves like $\log\epsilon$, while all the other terms have integrals that converge as $\epsilon\to 0$. So the only quantities that contribute to the finite part of $\mathscr{E}$ are the first two. For the first one we can use the expansion

$$\int_0^\infty dz\,\frac{2z^2+1}{\sqrt{z^2+1}}e^{-\epsilon z\Lambda_{\rm CP}}=\frac{2}{\epsilon^2\Lambda_{\rm CP}^2}+\frac{1}{2}+O(\epsilon), \tag{8}$$

while for the term proportional to ϵ^2 we use

$$\int_0^\infty(2z^2+1)e^{-\epsilon z}-\int_0^\infty\frac{2z^3+z}{\sqrt{z^2+1}}e^{-\epsilon\Lambda_{\rm CP}z}=\frac{1}{\Lambda_{\rm CP}\epsilon}-\frac{1}{3}+O(\epsilon) \tag{9}$$

plugging this two expansions back into Eq. (7) and subtracting the divergent part $N/\pi\epsilon^2$, one immediately gets that the leading term at large L of the energy

density is

$$\mathscr{E} \simeq -\frac{N}{4\pi}\Lambda_{\mathrm{CP}}^2. \tag{10}$$

Sub-leading behavior of the energy density may be calculated using a more refined approximation than the one we used here; however we can expect that those corrections to the result expressed above must be of at least $O(1/L^3)$ since, as we argued, the Lüscher term is inconsistent with a massive theory.

The total energy associated with the orientational moduli (translational moduli are not considered here) of $\mathbf{C}\mathrm{P}^{N-1}$ even with *DD* conditions seems to have a leading linear behavior in the length of the space interval as L becomes larger and larger.

Now in order to correctly interpret this results one needs to understand what this calculation actually gives. In terms of the energy of a M-V-M configurations, where the theory we are considering is a low-energy effective theory, the calculation we made represent a (small) correction to the classical string tension (that is also linear in L) due to vacuum fluctuation.

On the other hand if we regard the $\mathbf{C}\mathrm{P}^{N-1}$ model as a standing-alone "physical" two-dimensional QFT, the quantity that is fixed by experiments is the dynamically generated mass of the field n, but not the energy or energy density, whose renormalized value depends on the scheme used and it is therefore arbitrary.

Finally notice that those interpretations are possible, and consistent, with the fact that the energy density does not depend on the position x thus yielding a contribution that is proportional to L.

CONCLUSION

My conclusions have cost me some labor from the want of coincidence between accounts of the same occurrences by different eyewitnesses, arising sometimes from imperfect memory, sometimes from undue partiality for one side or the other.

— THUCYDIDES (*Peloponnesian War, I, 22)*

As we expected the study of the $\mathbf{C}\mathrm{P}^{N-1}$ model on a worldstrip is much more difficult than the model defined on the whole plane (even in the large N limit). This is of course due to the fact that the presence of the boundaries break the translational invariance through which the model can be solved in the infinite plane or under the assumption of periodic boundary conditions, leading to subtleties that are not present in the two dimensional model.

In spite of what has been suggested by Milekhin in [38], where the translational invariance has been assumed together with Dirichlet conditions, we have shown (with reference to the arguments proposed by Bolognesi, Konishi and Ohashi in [9]) that the system does not have two distinguished behavior (phases) one in which $m \sim \Lambda$ and $\sigma = 0$ (the confining phase) and the other in which $m = 0$ and $\sigma \sim \Lambda_{\mathrm{CP}}$ (the Higgs phase), rather it stays into just one phase that is the one that solves the full gap equations and that smoothly approaches the well known solution at $L \gg 1/\Lambda_{\mathrm{CP}}$.

The physical reason for this is the fact that at small L ($L \ll 1/\Lambda_{\mathrm{CP}}$) the quantum fluctuations have a wavelength that is constrained to be smaller than $1/\Lambda_{\mathrm{CP}}$ so that the system becomes essentially a quantum mechanical particle propagating in time.

Since the model we discussed emerges as a theory for the excitation of a M-V-M system that may have a role in explaining the mechanism of confinement in QCD, as we saw in section 1.2.4, it is most essential to have a good understanding of the energy stored in such a system. The central part of the thesis has then been devoted to the careful study of energy and energy density. We managed to write down an explicit representation of the energy density in terms of the eigenfunctions $f_n(x)$ and eigenvalues ω_n that describe the excitations of the fields. This expression for the energy is then shown to be consistent with the gap equations (in the sense that they can be derived by minimizing the energy density).

Then making use of the WKB approximation we show that the energy density can be made finite at fixed x.

Eventually we propose a proof of the fact that the energy density is constant in the interval $(0\,..\,L)$, and relying on this fact we find the dominant behavior of the energy at large L, showing that the orientatinal excitations contribute to the energy of the M-V-M complex with a term that is $E \simeq -N\Lambda_{\mathrm{CP}}/(4\pi)L$.

In conclusion, although this progresses have been made we still need to get a better understanding of some aspects of the problem; one of them is more pressing at them moment: We would like to understand if the result we have just mentioned on the energy is compatible with the regularization that we have used (a cutoff in the energy). Once this aspect will be perfectly clear we plan to expand the analysis we have done on the energy at large L to sub-leading terms and then eventually match the numerical results we currently have with analytical derivations to show the precise behavior of λ and σ as they approach the $L = \infty$ limit.

We are currently working on an article to clarify these questions we have raised.

APPENDIX A

INDEX TO NOTATIONS

If not otherwise stated letters that appears without any other kind of specification have the following meaning:

j, k, m, n	integer-valued arithmetic expression
x, y	real-valued arithmetic expression
z	complex-valued arithmetic expression
f	real-valued or complex-valued function
S	set
$\mathbf{Z}, \mathbf{R}, \mathbf{C}$	the set of integer, real and complex numbers

Formal symbolism	**Meaning**	**Where defined**
$\pi_1(X, x_0)$	fundamental group of X with basepoint x_0	1.1.2
$\pi_n(X, x_0)$	n-th homotopy group of X with basepoint x_0	1.1.2
$\deg F$	degree of the map F	1.1.2
$F^*\omega$	pullback of ω through F	
$Z(G)$	center of G	1.2.2.2
$B(x)$	borel transform of the power series $\sum_{n\geq 0} a_n x^n$	2
e_n^α	n-complex	2.1.2
$\chi(X)$	Euler characteristic of X: $\sum_n (-1)^n c_n$	2.1.2
$\mathbf{R}\mathrm{P}^n$	real projective space	2.2.1
$\mathbf{C}\mathrm{P}^n$	complex projective space	2.2.1
$[B]$	Iverson's notation for condition B	3.1
H_n	harmonic number	3.1.2.1
$I_\alpha(x)$	modified Bessel function of the first kind	3.1.2
$K_\alpha(x)$	modified Bessel function of the second kind	3.1.2
A^*	adjoint of A	
$[x\,..\,y]$	closed interval: $\{\, a \mid x \leq a \leq y \,\}$	
$(x\,..\,y)$	open interval: $\{\, a \mid x < a < y \,\}$	
∎	end of proof	

And to auoide the tediouse repetition of these woordes : is equalleto: I will sette as I doe often in woorke use, a paire of paralleles, or Gemowe lines of one lengthe, thus: ═══, bicause noe .2. thynges, can be moare equalle.

— ROBERT RECORDE, *The Whetstone of Witte* (1557)

APPENDIX B

BIBLIOGRAPHY

You will find it a very good practice always to verify your references, sir.

— MARTIN JOSEPH ROUTH, from J. W. Burgon, *Memoir of Dr. Routh* (1878)

HERE ARE THE REFERENCES to the works cited in the various chapters. They are written following the alphabetical order of the authors.

1 Milton Abramowitz, Irene Ann Stegun, *Handbook of Mathematical Functions*, United States Department of Commerce, National Bureau of Standards, ninth printing 1970.

2 Miguel Aguado, Manuel Asorey, "Theta-vacuum and large N limit in $\mathbf{C}\mathrm{P}^{N-1}$ sigma models," *Nuclear Physics B* **844** (2011), 243–265.

3 Damiano Anselmi, *Renormalization*, lectures notes from a course on renormalization, taught in Pisa in 2015, avaible online at the web page `http://renormalization.com`.

4 Manuel Asorey, David García-Alvarez and José María Muñoz-Castañeda, "Boundary effects in bosonic and fermionic field theories," *International Journal of Geometric Methods in Modern Physics* **12** (2015).

5 Roberto Auzzi, Stefano Bolognesi, Jarah Evslin, Kenichi Konishi, Alexei Yung, "Nonabelian superconductors: vortices and confinement in $\mathcal{N} = 2$ SQCD," *Nuclear Physics B* **673** (2003), 187–216.

6 Roberto Auzzi, Stefano Bolognesi, Jarah Evslin, Kenichi Konishi, Hitoshi Murayama, "Non-Abelian monopoles," *Nuclear Physics B* **701** (2004), 207–246.

7 Roberto Auzzi, Stefano Bolognesi, Jarah Evslin, Kenichi Konishi, "Non-Abelian monopoles and the vortices that confine them," *Nuclear Physics B* **686** (2004), 119–134.

8 Evgeny B. Bogomolny, "The stability of classical solutions," *Soviet Journal of Nuclear Physics* **24** (1976), 449–454.

9 Stefano Bolognesi, Kenichi Konishi, Keisuke Ohashi, "Large-N $CP(N-1)$ sigma model on a finite interval: physical boundary effects," arXiv:1604.05630 [hep-th].

10 Massimo Campostrini and Paolo Rossi, "Scaling properties of condensate in the $1/N$ expansion of lattice nonlinear σ-models," *Physics Letter B* **242** (1990), 81–88.

11 Massimo Campostrini and Paolo Rossi, "$\mathbf{C}\mathrm{P}^{N-1}$ models in the $1/N$ expansion," *Physical Review D* **45** (1992), 618–638; "Erratum: $\mathbf{C}\mathrm{P}^{N-1}$ models in the $1/N$ expansion," *Physical Review D* **46** (1992), 2741.

12 Massimo Campostrini and Paolo Rossi, "Dimensional regularization in the 1/N expansion," *International Journal of Modern Physics A* **7** (1992), 3265.

13 Massimo Campostrini, Paolo Rossi and Ettore Vicari, "Topological susceptibility and string tension in $\mathbf{C}\mathrm{P}^{N-1}$ models," *Physical Review B* **30** (1993), 830–833.

14 Massimo Campostrini, Paolo Rossi, and Ettore Vicari "Monte Carlo simulation of $\mathbf{C}\mathrm{P}^{N-1}$ models," *Physical Review D* **46** (1992), 2647.

15 Chandrasekhar Chatterjee, Kenichi Konishi, "Monopole-vortex complex at large distances and nonAbelian duality," *Journal of High Energy Physics* **39** (2014).

16 Giampaolo Cicogna, *Metodi matematici della Fisica*, Springer, 2008.

17 Mattia Cipriani, Daniele Dorigoni, Sven Bjarke Gudnason, Kenichi Konishi, and Alberto Michelini, "Non-Abelian monopole-vortex complex," *Physical Review D* **84** (2011).

18 Sidney Coleman, *Aspects of Symmetry*, Cambridge University Press, 1985.

19 Eugène Cremmer, Joël Scherk, "The supersymmetric non-linear σ-model in four dimensions and its coupling to supergravity," *Physics Letter B* **74** (1978), 341–343.

20 Alessandro D'Adda, Paolo Di Vecchia and Martin Lüscher, "A $1/n$ Expandable Series of Nonlinear Sigma Models with Instantons," *Nuclear Physics B* **146** (1978), 63–76.

21 Alessandro D'Adda, Paolo Di Vecchia and Martin Lüscher, "Confinement and chiral symmetry breaking in $\mathbf{C}\mathrm{P}^{n-1}$ models with quarks," *Nuclear Physics B* **152** (1979), 125–144.

22 Paolo Di Vecchia, Renato Musto, Francesco Nicodemi, Roberto Pettorino, Paolo Rossi, "The transition from the lattice to the continuum: $\mathbf{C}\mathrm{P}^{N-1}$ models at large N," *Nuclear Physics B* **235** (1984), 478–520.

23 Paul Adrien Maurice Dirac, "Quantized singularities in the electromagnetic field," *Proceedings of the Royal Society* **A133** (1931).

24 Gerald V. Dunne, Mithat Ünsal, "Resurgence and trans-series in Quantum Field Theory: the $\mathbf{C}\mathrm{P}^{N-1}$ model," *Journal of High Energy Physics* **170** (2012).

25 Gerald V. Dunne, Mithat Ünsal, "Continuity and Resurgence: Towards a continuum definition of the $\mathbf{C}\mathrm{P}^{N-1}$ model," *Physical Review D* **87** (2013), 025015.

26 Harald Eichenherr, "$SU(N)$ Invariant Nonlinear Sigma Models," *Nuclear Physics B* **146** (1978), 215–223.

27 Voislav L. Golo, Askold M. Perelomov, "A few remarks on chiral theories with sophisticated topology," *Letters in Mathematical Physics* **2** (1978), 477–482.

28 Amihay Hanany and David Tong, "Vortices, instantons and branes," *Journal Of High Energy Physics* **7** (2003), 37.

29 Allen Hatcher, *Algebraic Topology*, Cambridge University Press, 2001.

30 Donald E. Knuth, *The Art of computer Programming*, volume 1: *Fundamental Algorithms*. Addison-Wesley, 1968; third edition, 1997.

31 Kenichi Konishi, "The Magnetic Monopole Seventy-Five Years Later," *Lecture Notes in Physics* **737** (2008), 471-521.

32 Kenichi Konishi, "Physics of non-Abelian confinement and the dual gauge symmetry: Many faces of flavor symmetry," *Nuclear Physics B* **234** (2013), 333–336.

33 Kenichi Konishi, Giampiero Paffuti, *Quantum Mechanics: A New Introduction*, Oxford University Press, 2009.

34 John M. Lee, *Introduction to Smooth Manifolds*, Springer New York, 2012.

35 Stanley Mandelstam, "Vortices and quark confinement in non-abelian gauge theories," *Physics Letters B* **53** (1975), 476–478.

36 Stanley Mandelstam, "II. Vortices and quark confinement in non-Abelian gauge theories," *Physics Reports* **23** (1976), 245–249.

37 Nicholas Manton and Paul Sutcliffe, *Topological Solitons*, Cambridge University Press, 2004.

38 Aleksey Milekhin, "$CP(N-1)$ model on a finite interval in the large N limit," *Physical Review D* **86** (2012).

39 Charles W. Misner, Kip Stephen Thorne and John Archibald Wheeler, *Gravitation*, W.H. Freeman and Company, 1973.

40 Sergey Monin, Mikhail Shifman, Alexei Yung, "Non-Abelian String of a Finite Length," *Physical Review D* **92** (2015).

41 Philip M. Morse and Herman Feshbach, *Methods of Theoretical Physics*, part 1. McGraw-Hill, 1953.

42 Mikio Nakahara, *Geometry, Topology and Physics*, Tatlor & Francis; second edition, 2003.

43 Holger Bech Nielsen and Poul Olesen, "Vortex-line model for dual strings," *Nuclear Physics B* **61** (1973), 45–61.

44 Alexander Markovich Polyakov, "Particle spectrum in quantum field theory," *JETP Letters* **20** (1974).

45 Paolo Rossi, "On the effective Lagrangian of $\mathbf{C}\mathrm{P}^{N-1}$ models in the large N limit," submitted to *Physical Review D* (23.6.2016).

46 Walter Rudin, *Real and Complex Analysis*, McGraw-Hill; third edition, 1986.

47 Nathan Seiberg, Edward Witten, "Electric-magnetic duality, monopole condensation, and confinement in $\mathcal{N} = 2$ supersymmetric Yang-Mills theory," *Nuclear Physics B* **426** (1994), 19–52.

48 Mikhail Arkadyevich Shifman, *Advanced topics in Quantum Field Theory*, Cambridge University Press, 2012.

49 Franco Strocchi, *An Introduction to Non-Perturbative Foundations of Quantum Field Theory*, Oxford University Press, 2013.

50 Gerard 't Hooft, "A planar diagram theory for strong interactions," *Nuclear Physics B* **72** (1974), 461–473.

51 Gerard 't Hooft, "Magnetic monopoles in unified gauge theories," *Nuclear Physics B* **79** (1974).

52 Gerard 't Hooft, "Topology of the gauge condition and new confinement phases in non-abelian gauge theories," *Nuclear Physics B* **190** (1981), 455–478.

53 Steven Weinberg, *The Quantum Theory of Fields*, volume 1: *Foundations*. Cambridge University Press, 1995.

54 Steven Weinberg, *The Quantum Theory of Fields*, volume 2: *Modern Applications*. Cambridge University Press, 1995.

55 Steven Weinberg, *Lectures on Quantum Mechanics*. Cambridge University Press; second edition, 2015.

56 Edward Witten, "Instatons, the quark model, and the $1/N$ expansion," Nuclear Physics B **149** (1979), 285–320.

57 Edward Witten, "Baryons in the $1/N$ expansion," Nuclear Physics B **160** (1979), 57–115.